UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE MONTES

OBJECT-ORIENTED ANALYSIS OF REMOTE SENSING IMAGES FOR LAND COVER MAPPING: CONCEPTUAL FOUNDATIONS AND A SEGMENTATION METHOD TO DERIVE A BASELINE PARTITION FOR CLASSIFICATION[1]

TESIS DOCTORAL

GUILLERMO CASTILLA CASTELLANO

Ingeniero de Montes

2003

[1] To be cited as
Castilla, G. (2003) *Object-oriented analysis of Remote Sensing images for land cover mapping: conceptual foundations and a segmentation method to derive a baseline partition for classification*. Unpublished Ph.D. Thesis. Polythecnic University of Madrid. Madrid,Spain.

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE MONTES

Departamento de Economía y Gestión
de las Explotaciones e Industrias Forestales

OBJECT-ORIENTED ANALYSIS OF REMOTE SENSING IMAGES FOR LAND COVER MAPPING: CONCEPTUAL FOUNDATIONS AND A SEGMENTATION METHOD TO DERIVE A BASELINE PARTITION FOR CLASSIFICATION

Autor

Guillermo Castilla Castellano
Ingeniero de Montes

Directores

Joaquín Solana Gutiérrez
Doctor Ingeniero de Montes

Agustín Lobo Aleu
Doctor en Biología

2003

UNIVERSIDAD POLITECNICA DE MADRID

(D-18)

ACTA DE LECTURA Y DEFENSA DE TESIS DOCTORAL

ACTA NUMERO:

En Madrid, a seis de junio de 2003 y en la Escuela Técnica Superior de Ingenieros de Montes, se reúne y constituye el Tribunal encargado de juzgar la Tesis Doctoral de D. Guillermo Castilla Castellano

D.N.I. Arquitecto/Ingeniero/~~Licenciado~~ Ingeniero de Montes

Por La Universidad Politécnica de Madrid.

El Título de la Tesis es: Object-Oriented Analysis of Remote Sensing Images for Land Cover Mapping: Conceptual Foundations and a Segmentation Method to derive a Baseline Partition for Classification.

y ha sido dirigida por D. Joaquín Solana Gutiérrez D.N.I.

y D. Agustín Lobo Aleu D.N.I.

El Tutor del Doctorando ha sido D.: José Eugenio Martínez Falero

D.N.I. en el Programa de Doctorado de Economía y Gestión Forestal

del Departamento: Economía y Gestión Forestal

El Tribunal ha sido designado por el Excmo. Rector Magfco. de la Universidad Politécnica de Madrid con fecha 21 de marzo de 2003 y está integrado por los Doctores:

Presidente: D. Javier Martínez Millán

Vocales: D. Juan Ruiz de la Torre
D. Federico González Alonso
D. Carlos Pinilla Ruiz

Vocal Secretario: D. Fernando García Robredo

Suplentes: D. Mario Chica Olmo
D. José Luis González Rebollar

El Sr. Presidente significa que se ha dado cumplimiento a todos y cada uno de los trámites previstos en el Real Decreto 778/1998 de 30 de abril, y en el: "Reglamento de regulación de la solicitud, aceptación, desarrollo, lectura y defensa de la Tesis Doctoral" aprobado por el Claustro Universitario de la U.P.M.

El Tribunal declara abierta la sesión pública compareciendo ante él D. Guillermo Castilla Castellano.

Por el Vocal Secretario se da lectura al Artículo 10 del Real Decreto de 30 de abril de 1998.

Siendo las 12:09 horas y en aplicación de dichos preceptos, el Doctorando da comienzo a su actuación, consistente en la exposición de la labor preparatoria realizada, contenido de la tesis y conclusiones, haciendo especial mención de sus aportaciones originales.

Expuestas las opiniones de los miembros del Tribunal, sobre la Tesis leída y oídas las respuestas del Doctorando a las cuestiones y objeciones formuladas por aquellos, el Presidente invita a los Doctores presentes en la sala que formulen las cuestiones y objeciones que consideren oportunas.

Posteriormente el Tribunal invita al Doctorando y al público asistente a que se ausenten de la sala y, reunido en sesión privada comienza su deliberación, pára lo cual todos y cada uno de los Sres., Vocales, y el Sr. Presidente exponen su criterio con respecto a la actuación del Doctorando en defensa de su Tesis Doctoral.

Concluye la deliberación del Tribunal y previa votación en sesión secreta a las 13:46 horas se acuerda otorgar a la Tesis Doctoral la calificación de:*

SOBRESALIENTE CUM LAUDE

Calificación otorgada con (5) votos.

OBSERVACIONES:

EL PRESIDENTE,

Fdo. Javier Martínez Millán

VOCAL,

Fdo. Juan Ruiz de la Torre

EL VOCAL SECRETARIO,

Fdo. Fernando García Robredo

VOCAL,

Fdo. Federico González Alonso

* Sobresaliente cum laude
Sobresaliente
Notable
Aprobado
No Apto.

VOCAL,

Fdo. Carlos Pinilla Ruiz

TABLE OF CONTENTS

PREFACE

"Individuals who break through by proposing a new paradigm are almost always...either very young or very new to the field whose paradigm they change....These are the men who, being little committed by prior practice to the traditional rules of normal science, are particularly likely to see that those rules no longer define a playable game and to conceive another set that can replace them."
-- Thomas S. Kuhn in *The Structure of Scientific Revolutions* (1962).

Although inspiring, the above quotation is by no means applicable to myself. However, it keeps some resemblance with the circumstances that gave birth to this thesis. After completing in 1990 my studies in Forest Engineering at the Polytechnic University of Madrid, I began the doctoral courses hoping that I would get some financial support to undertake a Ph.D. Thesis. Since unfortunately this was not the case, I completed the courses and took another path that got myself involved in development aid for several years. During all that period I had with me this unsatisfied academic zeal. Vg. while in Mozambique, I applied for a Ph D. program at CATIE in Costa Rica and even was accepted, but being a national of a developed country, I was suggested to get my own funding...

The path to this thesis began in Málaga in Christmas 1997. I was in a dinner with some old university fellows when I received a proposal from the Forest Map of Spain (MFE) team, with whom I had collaborated in the past, to apply for an ESA/CDTI grant: I would finally have data and time to write my Ph.D. Thesis and they would get in turn a method to update the Map with satellite imagery. I applied for it and some months later, while working in Guatemala, I was informed that I had been granted by the Spanish Ministry of Education to stay for two years as a trainee in ESRIN, the European Space Agency (ESA) Earth Observation data handling and exploitation centre, located in Frascati, Italy. I quitted my overseas job and started my stay in ESRIN in December 1998.

By the time I arrived there, I hardly had some bare notions about Remote Sensing, so it took me several months to familiarize with this field. But as soon as I got my first Landsat image and could overlay the MFE coverage on it, in observing that most polygons were distinguishable from their neighbours, I realized a somehow naïve thought that became the motto of this research: '*If the eye can see the difference, the problem is not in the data but in the analysis*'. Such trivial conclusion, derived from the poor results I obtained both from supervised and unsupervised classifications, led me to embark on an exciting inquiry that for the time being has ended up in the present document.

Unfortunately, my ESRIN traineeship came to an end before I could manage to put all the pieces of knowledge I had gathered into a workable scheme. I retired to my hometown Granada in order to carry on with the thesis, with the only sponsorship of my parents, who generously provided me with food and shelter, and of some temporary jobs I got with the help of friends and colleagues, which allowed me to retain my self-esteem. Albeit such voluntary seclusion has been hard to bear, it gave me a total intellectual freedom without which I probably would have never written a thesis like this. However, I was never in isolation during that period. I was connected to the world through a thin ADSL wire that allowed me to keep in touch with colleagues and to search the internet for the most diverse documents. Indeed, this thesis is a product of the internet era, for it gave me the wings I needed in order to collect the nectar of scattered ideas. Having dissolved them into a loosely coupled whole –this thesis, now it is time for others to transform it into a richer nourishment for the progress of this field.

Granada, Spain, December 2002

ACKNOWLEDGMENTS

During the long four years that took this thesis to complete, I received the help and support of many people and institutions. First of all, I am particularly indebted to the MFE team, especially to Juan Ignacio García Viñas and Professor Juan Ruiz de la Torre. The former embarked me in this exciting enterprise and gladly helped me in numerous tasks, including a visit to the study sites, the arrangement of some presentations, and networking with some individual and institutions. The latter offered me a generously paid consultancy job when, some months after coming back to Spain from Italy and running out of money, I was about to give up and resume my activity overseas as an aid worker. They also supported my application to the ESA traineeship, which turned out to be decisive for my acceptance.

I am also thankful to the Spanish Ministry of Education and the 'Centro para el Desarrollo Tecnológico Industrial' (CDTI), particularly to Mr. Manuel Serrano, for the concession of a grant from the 'subprograma de becas de especialización en organismos internacionales', which allowed me to stay for two years at ESRIN as a trainee. There at Frascati there is a great bunch of people to whom I am indebted. I whish to thank my successive tutors at ESRIN, Marc Gorman, Luigi Fusco and Olivier Arino for providing me with all the data and facilities I needed. Mr. Juerg Lichtenegger, although not his duty, acted as a mentor who guided the first steps of this research and not only, he was always available and even lent me his personal laptop when I suspended my traineeship during a month to travel to Mozambique to help a local NGO after the flooding of March 2000. Massimo Barbieri and Simone Paoloni kindly helped me in many occasions, mostly related to the processing of images requiring a specialised software, but not only. Maria Cristina Guarrera and Monica Rollán helped me with the gathering of papers even after having returned to Spain. And many other friends gave me a hand and counsel perhaps more often than they wished, to mention a few: Daniel Esteban, Marcello Mariucci, Nick Walker, Josep Closa, Lucio Castellano, Carlos Linares, Diego Fernández, and many others...

Out of ESRIN, I wish to express my gratitude to Dr. Håkan Olsson, who gave me the opportunity to visit Sweden and meet his team at the SLU RS lab in Umeå. I am also thankful to Paul Smits, Olle Hagner, Mario Chica-Olmo and Javier Martínez-Millán for helpful comments.

Many thanks to my supervisors, Dr. Joaquin Solana, who gently urged me to stop digressing and to write this otherwise never ending story, bore my long explanations, cleared the bureaucratic laberynth, and even helped me to get a temporary academic job; and Dr. Agustin Lobo, for having stood up all this period even after the refusal of the funding proposal that was supposed to back up our relationship, for his innumerable sharp comments, for proofreading drafts, and for bearing my stubbornness and reluctance to be guided. I am also indebted to my tutor during my university years, Dr. Eugenio Martínez-Falero, for many things, but particularly for having managed to get for me the extensions I should have requested while I was working overseas in order to maintain open the possibility of completing the Ph.D. degree.

Special thanks to Reyes Ruiz for his friendship, encouragement and support, without which the Spanish period of this research would had been much harder. Other best friends have being cheering me up, specially in that final 10% that takes 90% of the time, among them Ramon Santiago. Ending up this list with the most beloved persons, I would like to thank my parents and siblings, for their loving and material support, and for their enormous patience. I still remember telling them shortly after coming back from Italy that it would only take me a few months... My final thanks go to my girlfriend, Victoria, who in addition to love and care has given me the emotional balance I needed in order to face this personal challenge.

A mis padres, sin cuyo generoso apoyo esta tesis jamás hubiese visto la luz

“ANALISIS ORIENTADO A OBJETOS DE IMÁGENES DE TELEDETECCIÓN PARA CARTOGRAFIA FORESTAL: BASES CONCEPTUALES Y UN METODO DE SEGMENTACION PARA OBTENER UNA PARTICION INICIAL PARA LA CLASIFICACION”

RESUMEN

La toma de conciencia de la opinión pública sobre el medio ambiente, cuyo inicio se puede situar después de la Conferencia de Estocolmo de 1972 y su madurez tras la Cumbre de Río de 1992, ha dado lugar a la creación de ministerios y agencias regionales dedicados a su gestión y protección en la mayoría de los países. Estas instituciones precisan para desempeñar eficazmente su labor de información detallada y al día sobre aspectos relevantes del territorio que controlan. En consecuencia, la demanda de información geográfica ha crecido sustancialmente en las últimas décadas, a un ritmo quizá sólo superior al del desarrollo de las tecnologías que la hacen posible y accesible a los ciudadanos.

Con base en esa información geográfica, las autoridades toman decisiones, y a través de ella el público se forma una opinión que a su vez influye en éstas. Por tanto la calidad de la información, en términos de extensión, detalle y frescura, es un requisito ineludible para una sociedad pudiente preocupada por su medio natural. Y en un país como España, donde cada año se construyen numerosas infraestructuras, se producen cambios de uso de suelo de rústico a urbano y de agrícola o improductivo a forestal, y donde las cubiertas vegetales pueden cambiar drásticamente a causa de los incendios, la necesidad de disponer de información geográfica actualizada es aún más patente.

De entre esa información, uno de los temas más importantes es la distribución espacial de los diferentes tipos de cubierta existentes, pues en función de su naturaleza y estado se ordenan las actuaciones y normativas en cada zona concreta. En las áreas naturales (entendidas como no agrícolas ni urbanas), la vegetación –o su ausencia- es el factor primordial que define no sólo los tipos de cubierta, sino las otras características de los ecosistemas que constituye. En España se denominan forestales a todas las zonas cubiertas por vegetación natural aunque consistan en formaciones no boscosas, por lo que en adelante se hablará de cartografía y de mapas forestales, cuya actualización es el problema de fondo que esta tesis aborda.

La información sobre las zonas forestales tradicionalmente se ha obtenido por medio de recorridos de campo e inventarios que recogen datos que luego son resumidos en una base de datos geográfica. Hasta hace unos pocos años, esa base de datos se presentaba de forma visual y escueta como un dibujo simbólico en una hoja grande de papel: el mapa forestal. Los mapas tratan de hacer observables a una escala háptica (esto es, distinguibles a simple vista a una distancia inferior a la que alcanza la mano) los aspectos de interés de un territorio relativamente extenso. Lo que vemos por ejemplo en un mapa topográfico 1:50.000 de una zona dada sosteniéndolo con los brazos extendidos se podría aproximar a lo que veríamos a través de un marco de las mismas dimensiones colocado horizontalmente y sostenido de igual forma a unos 25.000 m. de altitud sobre esa zona, con la diferencia de que en el mapa los elementos de interés aparecen resaltados y/o exagerados, como p.ej. las carreteras.

Los mapas forestales usan normalmente como fondo un mapa topográfico sobre el que se añaden una serie de recintos contiguos o *teselas* que corresponden a zonas relativamente homogéneas respecto al tipo de cubierta vegetal. Sobre cada tesela se dan una serie de informaciones visuales a través de colores, tramas o sobrecargas geométricas, símbolos y letras. Generalmente se acompaña el mapa de una memoria en la que se exponen otra serie de informaciones que por su carácter excesivamente detallado (o bien general) no pueden ser representadas en el mapa. La escala empleada depende del objetivo del mapa, de las restricciones presupuestarias , y de la base cartográfica disponible, que típicamente oscila de 1:10.000 (mapas a nivel local) a 1:1.000.000 (a nivel nacional).

El desarrollo de la fotogrametría a partir de la Primera Guerra Mundial supuso un notable avance para la cartografía forestal, que dejó de basarse en croquis y mediciones sobre el terreno para apoyarse en fotos aéreas, técnica aún vigente hoy en día. Las teselas se delinean manualmente sobre las fotos y posteriormente se pasan al mapa con la ayuda de un restituidor. La *fotointerpretación* consiste en la delimitación e identificación de regiones homogéneas en la imagen, y se basa en las diferencias visuales que produce cada tipo de cubierta. Los caracteres que se usan a este fin son el color o el tono, la textura, el tamaño, forma y patrón de distribución de los objetos, y el contexto. Todos esos elementos tomados en su conjunto permiten al fotointérprete establecer un diagnóstico sobre el tipo de cubierta presente en cada tesela, lo que posibilita cartografiar la vegetación con relativamente poco trabajo de campo. Este consiste normalmente en la inspección de algunas teselas según un determinado muestreo, el cual se realiza normalmente en dos etapas: una primera para la propia elaboración del mapa, y otra posterior para estimar su precisión (frecuencia de errores de clasificación y/o delineación).

El proceso de producción de cartografía forestal descrito arriba es demasiado lento y costoso como para permitir una frecuencia de actualización similar a la de los cambios en el territorio, siendo el periodo entre renovaciones tipicamente superior a diez años. Este método requiere una cantidad considerable de personal especializado, y posiblemente un vuelo fotogramétrico *ad hoc*; además presenta el problema de la subjetividad del fotointérprete a la hora de trazar los bordes de las teselas, lo que en futuras revisiones muchas veces dará lugar a correcciones aun sin haberse producido cambios. Sin embargo, los avances en las dos últimas décadas en la informática, que han disminuido drasticamente el coste de almacenamiento de datos e incrementado espectacularmente la capacidad de procesamiento; en telecomunicaciones, que permiten un intercambio masivo e instantáneo de información entre cualquier parte del mundo; y en teledetección, que han aumentado continuamente la resolución espacial, espectral y temporal de los datos recogidos por los satélites de observación de la Tierra, han cambiado totalmente el panorama.

Hoy en día los mapas forestales ya no son unicamente una serie de informaciones visuales presentadas en papel. Están integrados en un Sistema informatizado de Información Geográfica (SIG) que almacena y explota grandes bases de datos geográficos estructuradas en capas temáticas, las cuales pueden ser combinadas facilmente para adaptarse a las necesidades del usuario. Los SIG posibilitan la cartografía automática de los resultados de un determinado análisis estadístico, y la reproducción de la información, bien en la pantalla o impresa, a cualquier escala y con diferentes presentaciones. La expansión comercial del software de SIG, cada vez más versátil y económico, unido a la creciente difusión de internet, permiten en la actualidad a los ciudadanos acceder con facilidad a toda clase de informaciones sobre su territorio, con lo que la demanda de este tipo de información está creciendo exponencialmente.

Los avances mencionados han llevado a un mayor uso de métodos automáticos de análisis de imágenes de satélite para la elaboración de cartografía temática. Sin embargo, la calidad de los mapas resultantes está por debajo de los estándares exigidos por las instituciones que los encargan. En consecuencia, la mayoría de los proyectos de cartografía aún se apoyan en cierta medida en la fotointerpretación. No obstante, ésta técnica no sólo es menos eficiente, sino que conlleva un alto grado de subjetividad. Por tanto, las conclusiones sobre el cambio de tipo de cubierta derivadas de la comparación de sucesivas actualizaciones realizadas por este método son poco fiables. Considerando el mayor énfasis que actualmente se da al seguimiento de los cambios en el territorio, hay una mayor necesidad de mejorar y agilizar el modelado del paisaje. Para ello se requieren métodos automáticos, sólidamente fundados y de mayor precisión, que constituyan la base operativa sobre la que mantener actualizada la información geográfica que orienta las actuaciones de las agencias territoriales de medio ambiente.

Por otro lado, el enfoque comúnmente usado para analizar las imágenes de satélite con fines cartográficos da lugar a resultados insatisfactorios debido principalmente a que unicamente utiliza los patrones espectrales de los píxeles, ignorando casi por completo la estructura espacial de la imagen. Este enfoque (denominado en esta tesis 'espectrométrico') se basa en la discriminación de firmas espectrales, las cuales están normalmente constituidas por los valores que adopta cada píxel en cada una de las bandas que constituyen la imagen multiespectral, las cuales se forman según se detalla a continuación. El sensor adquiere datos que son agrupados espacialmente en una matriz o *ráster* en la que cada celdilla corresponde a una medición de una señal eléctrica que es función de la cantidad de radiación recibida en esa posición y momento. La medición es discretizada por un convertidor analógico-digital a una escala finita o *rango dinámico* (de 0 a 255 para la mayoría de los sensores ópticos), y el valor resultante es introducido en esa celdilla en forma de Número Digital (DN). La radiación incidente es separada antes de alcanzar los detectores del sensor por medio de un sistema de prismas y filtros, de forma que cada banda de una imagen multiespectral corresponde a la radiación capturada en un intervalo particular del espectro electromagnético.

Los valores de cada celdilla, representados a lo largo del espectro, se pueden interpolar dando lugar a una curva o firma espectral, que aunque más grosera tiene una cierta similitud con la que se obtiene de los espectrómetros de sobremesa, de ahí que cada píxel se considere como una muestra individual. Estas firmas se pueden también representar como puntos de un espacio multivariante en el que cada eje ortogonal se refiere a una banda y está constituido por el conjunto ordenado de valores del rango dinámico. La clasificación espectrométrica de las imágenes consiste por tanto en delimitar las regiones de ese espacio asociadas a cada clase. Las clases deben satisfacer los siguientes requisitos:

- Exhaustividad: debe haber una clase que asignar a cada píxel de la imagen.

- Separatividad: las clases deben ser separables en el espacio multivariante con el clasificador empleado.
- Utilidad: las clases deben cubrir las necesidades de información del usuario.

El requisito de separabilidad implica que las firmas de clases diferentes estén relativamente distantes las unas de las otras, de forma que su grado de solape sea despreciable. Sin embargo para que esto se cumpla, cada *cluster* (nube de puntos) del espacio multivariante debe contener una clase mayoritaria. Por otro lado, la cuadrícula de terreno sobre la cual el sensor realiza la medida de cada píxel debe ser suficientemente grande como para incluir los elementos que producen la firma espectral típica de cada clase. Dicho de otro modo, el tamaño de la cuadrícula debe ser tal que un observador situado en el centro de ella tenga suficientes elementos de juicio como para asignar la clase correcta restringiendo la observación al interior de la cuadrícula. El tamaño mínimo necesario para realizar una clasificación correcta sobre el terreno recibe el nombre de resolución espacial de la clasificación. Por tanto, una premisa básica del enfoque espectrométrico es que la extensión de la cuadrícula sobre la que se extrae la muestra (es decir, el tamaño del píxel) supere esa resolución.

Ahora bien, cuanto mayor sea el tamaño de la cuadrícula, mayor será el porcentaje de píxeles "mezclados", es decir, píxeles incluyen bordes entre teselas. Como la firma espectral de éstos es una mezcla de las firmas típicas de las clases de esas teselas, su posición en el espacio multivariante corresponderá a las zonas que separan clusters de clases diferentes. Sin embargo, el propio concepto de cluster requiere que éste esté separado de otros por una discontinuidad, eso es, por una región del espacio multivariante casi vacía. Por tanto la premisa del tamaño suficiente y del solape despreciable no pueden ser satisfechas simultáneamente cuando las unidades analizadas son píxeles individuales, a no ser que se estudie un territorio relativamente simple (como un paisaje agrícola con grandes campos de cultivo) con unas clases de cubierta muy generales. A pesar de esto, el enfoque basado en píxeles sigue siendo el paradigma comunmente aceptado en esta disciplina.

Para entender el porqué de esta situación, nos tenemos que remontar a los orígenes de la teledetección espacial allá por los años setenta. El tamaño de píxel con que ésta comenzó (80 m) era compatible con la resolución espacial de la mayor parte de las clasificaciones. A esa escala, era más natural considerar las clases de cubierta como materiales homogéneos distribuidos sobre el territorio en parcelas mayores que un píxel, por tanto era razonable asimilar cada píxel individual a una muestra introducida en un espectrómetro que puede ser analizada por separado. Conforme el progreso técnico permitió mayores resoluciones, la variabilidad radiométrica de las clases aumentó, por tanto se hizo necesario por un lado incorporar al análisis alguna característica espacial como la textura que pudiera paliar esta heterogeneidad, y por otro realizar un pre y/o un post-procesamiento basado en filtros que redujese la inconsistencia espacial de las imágenes clasificadas.

La aparición a principios de este siglo de satélites civiles de muy alta resolución (< 5m.) ha puesto de manifiesto las deficiencias del enfoque espectrométrico cuando no se cumple la premisa del tamaño. Además, la equiparación de las clases de cubierta a tipos de materiales homogéneos permite que cualquier parte arbitrariamente delimitada dentro de una tesela del mapa siga siendo un referente del concepto definido por su etiqueta. Esta posibilidad es incongruente con el modelo jerárquico del paisaje cada vez más aceptado en Ecología del Paisaje, que asume que la homogeneidad depende de la escala de observación y en cualquier caso es más semántica que biofísica, y que por tanto los paisajes son intrínsecamente heterogéneos y están compuestos de unidades ('patches') que funcionan simultaneamente como un todo diferente de lo que les rodea y como partes de un todo mayor. Por tanto se hace necesario un nuevo enfoque (orientado a objetos) que sea compatible con este modelo y en el que las unidades básicas del análisis sean delimitadas de acuerdo a la variación espacial del fenómeno estudiado. Esta tesis pretende contribuir a este cambio de paradigma en teledetección, y sus objetivos concretos son:

1) Poner de relieve las deficiencias del enfoque tradicionalmente empleado en la clasificación de imágenes de satélite.
2) Sentar las bases conceptuales de un enfoque alternativo basado en zonas básicas cartografiables.
3) Desarrollar e implementar una versión demostrativa de un método automático que convierte una imagen multiespectral en una capa vectorial formada por esas zonas.

El modelo jerárquico concibe el paisaje como un mosaico de unidades funcionales-estructurales anidadas jerárquicamente, de forma que cada unidad se puede considerar compuesta de subunidades que interactúan más entre ellas que con subunidades de unidades vecinas, con lo que cada unidad forma un todo más o menos integrado, esto es, un objeto. A cada nivel superior de la jerarquía le corresponden objetos cada vez mayores: árbol, bosquete o rodal, bosque. Para cada nivel jerárquico se puede establecer un umbral de tamaño (la unidad mínima cartografiable) por debajo del cual se asume que no existen –o no interesan- objetos de ese nivel. Bajo este prisma, cada clase de cubierta es un concepto geográfico que se refiere a una serie de objetos de un nivel específico de esa jerarquía que tienen una estructura y funcionamiento similares. Por tanto una región arbitrariamente delimitada dentro de esos objetos no puede ser un referente de ese concepto, pues le faltan la unidad del todo y la diferencia con el resto (una parcela de inventariación dentro de un bosque no es un bosque).

El hecho de que muchos de estos objetos geográficos tengan bordes difíciles de delimitar no implica que su interior sea ontológicamente dependiente del mapa, más bien son los bordes que los separan las creaciones humanas, y es la selección de los criterios objetivos para su delimitación lo que genera su dependencia del analista. Así, la multiplicidad de alternativas que existen a la hora de dibujar el límite de un bosque no refleja más que la vaguedad del concepto *bosque*, y esa vaguedad solo puede ser reducida añadiendo a la definición del diccionario enunciados más precisos sobre lo que es un bosque bajo la óptica del mapa, que sean cuantificables con los medios disponibles para su elaboración. Otro problema relacionado es cualquier bosque, siendo un objeto complejo, contiene partes (p.ej. calveros) que si son aisladas de su entorno no pueden ser reconocidas como partes del bosque, por lo que el enfoque espectrométrico, que supone que el bosque es homogéneo en toda su extensión, presenta serias deficiencias a la hora de abordar la heterogeneidad intrínseca de los paisajes. Se hace por tanto necesario un nuevo enfoque que sea más acorde con el modo natural de interpretar el paisaje que tenemos los humanos, consistente en dividirlo en una serie de entidades discretas u objetos que son referentes de la serie de conceptos que usamos para dar sentido a lo que vemos.

El enfoque alternativo que se sigue en esta tesis está basado en el análisis orientado a objetos, que trata de modelar el campo de estudio usando objetos que son *instancias* (ejemplos particulares) de clases, y que es especialmente adecuado para analizar el paisaje bajo el citado modelo jerárquico. Los objetos se agrupan en clases organizadas jerarquicamente que permiten a las clases inferiores *heredar* propiedades de las clases superiores de las que proceden, y esta estructura da lugar a una jerarquía de objetos en subobjetos y superobjetos que a su vez permite la *encapsulación* (ocultación) de la información. A la vista del fracaso del enfoque espectrométrico cuando se aplica a imágenes de los nuevos sensores de más alta resolución, el análisis orientado a objetos está adquiriendo cada vez más auge en teledetección, aunque de momento carece de una base teórica sólida específica a este campo y a su aplicación a la cartografía del paisaje. Esta tesis intenta sentar las bases conceptuales de este enfoque, proporcionando además un método automático de segmentación para obtener una partición inicial de la escena que sirva de base a la clasificación.

La estrategia que se propone es producir, basándose en la estructura espacial de las imágenes, una partición de estas en la que cada región puede considerarse relativamente homogénea y diferente de sus vecinas, y que además supera (aunque no por mucho) el tamaño de la unidad mínima cartografiable. Estas regiones son las unidades básicas de la clasificación, sobre las cuales se pueden definir una serie de atributos espaciales (forma, tamaño, orientación), estructurales (disposición interna, tono, textura y contraste entre las diferentes partes que las componen) y contextuales (relaciones con regiones vecinas) que no son aplicables a píxeles individuales. Cada región se asume corresponde a un rodal que tras la clasificación será agregado junto a otros rodales vecinos en una región mayor que en conjunto pueda verse como una instancia de un cierto tipo de objetos que más tarde son representados en el mapa mediante teselas de una clase particular.

Esta estrategia se basa en un modelo en tres niveles del territorio, en el que el último nivel es el mapa forestal en sí. El primer nivel está constituido por objetos solidarios a la superficie terrestre y no observables a las escalas de esos mapas, como árboles y edificios. El segundo nivel está formado por una serie de campos geográficos (variables regionalizadas continuas), que en la versión idealista del modelo corresponden una serie de atributos cuantitativos, uno por campo, derivados de la medición (en parcelas circulares centradas en cada punto del territorio) de alguna característica relacionada con las propiedades y/o distribución espacial de los objetos del primer nivel. En la versión realista del modelo no es posible obtener los anteriores campos por motivos económicos, y en su sustitución se recurre a otros campos geográficos, indirectamente relacionados con los atributos biofísicos usados para clasificar el paisaje, que son el conjunto de ortoimágenes de satélite disponibles. Finalmente, en el tercer nivel, las diferencias cuantitativas reflejadas por los campos son transformadas en diferencias cualitativas entre objetos, esto es, se pasa de una representación numérica del territorio, fácilmente manipulable por un ordenador, a una representación simbólica, mucho más asimilable y manejable por una persona.

El paso del segundo al tercer nivel se realiza en primera instancia por un proceso de morfogénesis basado en la teoría de catástrofes, que da lugar a la delimitación de una serie de regiones primarias disjuntas que corresponden al área de influencia de cada atractor local (punto de mínima variación en una determinada región del campo). Las regiones primarias adyacentes se unen entre sí según el parecido de sus valores medios en los atributos considerados, y esta agregación tiene lugar hasta que las regiones resultantes alcanzan el tamaño mínimo que se les supone a los objetos que se pretenden clasificar. A partir de ese punto comienza la clasificación, en la que el usuario aporta su conocimiento sobre cada tipo de objeto que quiere resaltar en el territorio, y lo expresa p.ej. a través del rango de valores admisibles para cada clase y atributo, y adicionalmente, como un conjunto de reglas relacionales que pueden conducir al cambio de significado (etiqueta de clase) de un objeto según su contexto. Un posible método de clasificación es propuesto pero no desarrollado en la tesis.

En la versión realista del modelo, el proceso de morfogénesis se simula mediante un filtrado difusivo no lineal (que elimina la textura y respeta los bordes) de la ortoimagen de satélite, seguido de una transformación geodésica en microcuencas ('watersheds') de la imagen conjunta de magnitud del gradiente, transformación que equipara esta última imagen con un modelo digital del terreno, y que delimita el área de influencia de cada mínimo local de gradiente. Para crear la imagen de gradiente, se define previamente una medida de la disimilitud radiométrica entre píxeles adyacentes de la ortoimagen multiespectral filtrada, de forma que los números digitales asociados a los píxeles de la imagen de gradiente representan el valor absoluto de la máxima variación en similitud existente en el entorno de cada píxel de la ortoimagen. Las 'microcuencas' son a continuación agregadas por adyacencia siguiendo un orden definido por su similaridad radiométrica, hasta que las regiones resultantes superan todas el tamaño de unidad mínima cartografiable.

En cada iteración, se comienza uniendo los pares de regiones adyacentes que presentan la diferencia más baja, permitiéndose una única unión por región e iteración, e impidiéndose la unión cuando alguna de las dos regiones tiene un vecino más similar que el que se está evaluando o cuando alguna de las regiones colindantes ha sido ya unida a otra en la iteración en curso (ya que el nuevo agregado podría ser el vecino más similar de una de las dos regiones). De está manera, se forman primero regiones más homogéneas, a las que progresivamente se van incorporando regiones más pequeñas y de más alto contraste con su alrededor. Así, la imagen 'labelizada' (en la que cada píxel tiene como DN el identificador numérico de la región a que pertenece, salvo los píxeles de borde, que tienen su DN=0) que representa el estado de la partición se va simplificando progresivamente durante la segmentación, hasta que todas las regiones superan el tamaño mínimo especificado.

Además de constituir las unidades básicas propuestas para efectuar una clasificación orientada a objetos, las regiones resultantes de la segmentación forman un retículo que podría ser la base de un procedimiento de fotointerpretación asistida, en la que el/la intérprete tan sólo tiene que seleccionar y eliminar (con clics de ratón) los arcos irrelevantes, es decir aquellos que considere que separan regiones que para él/ella son lo mismo. Otra aplicación inmediata del método es la verificación/actualización de mapas forestales mediante la detección de zonas incongruentes (regiones básicas que presentan una apariencia diferente de las demás regiones circunscritas en la misma tesela).

El modelo en tres niveles del territorio y el proceso para construirlo se basan en tres hipótesis. La primera, que permite pasar de la versión idealista a la realista, asume que la variación espacial conjunta de los campos geográficos ideales asociados a los atributos biofísicos de la cubierta coincide en su mayor parte con la variación conjunta de luminancia de las bandas que componen la ortoimagen multiespectral. Dicho de otra forma, la *hipótesis de coincidencia* presupone que los bordes que aparecen en la imagen de gradiente conllevan un cambio significativo en alguno de los atributos biofísicos relevantes para la clasificación de la cubierta de las regiones separadas por esos bordes. El siguiente paso para construir el modelo es reconocer que el único modo de abordar la heterogeneidad jerárquica del paisaje consiste en establecer un nivel de generalización por debajo del cual carece de interés conocer la estructura interna de las partes que componen las unidades que se pretenden identificar. Este reconocimiento se materializa en la *hipótesis de tamaño*: para que un rodal o mancha tenga interés y por tanto cabida en el modelo, este debe superar una cierta extensión mínima, que en principio se puede asimilar a la de la unidad mínima cartografiable. Por tanto, aquellas regiones primarias (derivadas de la morfogénesis) que no lleguen a ese tamaño tendrán que ser agregadas. Este último proceso es guiado por la *hipótesis de correspondencia* entre similitud radiométrica (espectral) y similitud semántica (taxonómica) en regiones adyacentes, que asume que si dos regiones vecinas tienen la misma apariencia en la imagen, entonces es muy probable que correspondan al mismo tipo de cubierta.

El método para obtener la partición inicial del territorio, que sirve de punto de partida para la clasificación, se ha implementado en un lenguaje de programación especialmente adecuado para procesar imágenes digitales (IDL) y se ha probado tanto en imágenes multiespectrales como en ortofotos pancromáticas aéreas. Los resultados preliminares se adaptan razonablemente bien a la estructura espacial de la imagen. Cada región definida se percibe como una unidad que presenta una cierta coherencia interna que además se ve diferente de lo que la rodea. Sin embargo, los resultados también revelan una serie de problemas:

Poca consistencia, en áreas de bajo contraste, de los bordes resultantes de aplicar el método con diferentes parámetros del filtrado inicial. Aunque esta circunstancia era previsible, indica una falta de robustez del método en ausencia de un contraste claro entre regiones diferentes.
Dependencia del resultado de la frecuencia de actualización física de la imagen 'labelizada', que no se efectúa en todas las iteraciones por razones de economía de cómputo. Este problema impide de momento que el método se pueda aplicar a imágenes grandes (mayores de 4 Mpixels) que requieran ser subdivididas en imágenes más pequeñas para su procesamiento, ya que se producirían incongruencias a la hora de casar las subimágenes.
Efectos fractales en los bordes resultantes, que se evidencian por que la longitud del perímetro de una región cualquiera crece indefinidamente al disminuir el tamaño del píxel, es decir, que los bordes de las regiones son tanto más complejos cuanto más alta es la resolución. Estos efectos pueden dificultar la vectorización de la partición.

El trabajo futuro se puede centrar en tres áreas, relacionadas respectivamente con el modelo conceptual del paisaje, el método de segmentación, y la clasificación de las regiones resultantes de ésta. El modelo en tres niveles es susceptible de una definición más formal por medio de un conjunto lógico-matemático de axiomas, definiciones, propiedades y relaciones. Las deficiencias referidas del método de segmentación se pueden solventar con relativa facilidad. Por un lado, la significación radiométrica de los bordes de la partición inicial en microcuencas se puede evaluar a partir de su prominencia geodésica, y partir de ella se puede simplificar la partición inicial para que no incluya bordes débiles. Los problemas de actualización de la partición durante las iteraciones se pueden remediar si se convierte la partición de microcuencas a una capa vectorial de tipo polilínea en la que cada arco tiene un identificador relacionado con el de las regiones que separa además de una serie de atributos que registren el valor medio de los píxeles del arco en cada una de las bandas. Y los efectos fractales pueden ser mitigados si se ordenan los vértices en una jerarquía de escala similar a la de las regiones clasificadas, de forma en cada escala de visualización, tanto el número de regiones representadas como la densidad de bordes permanecen más o menos constantes.

Por último, la clasificación de las regiones de partida requiere por un lado definir una serie de atributos sustitutivos de los biofísicos sobre los que apoyarla, y por otro, una serie de reglas heurísticas para efectuarla. Una posibilidad es aplicar el método ELECTRE de ayuda a la toma de decisiones multicriterio. Primero, se estima el rango de valores admisibles de cada atributo en cada clase, para lo cual se requiere información fidedigna de inspecciones sobre el terreno o de fotografía aérea detallada, contrastada por la información del mapa que se pretende actualizar. Después, para cada región *i* se compara el valor que toman en ella los atributos con el rango de valores de cada clase *c*. Entonces el conjunto de atributos es dividido en tres partes, según el valor de estos sea concordante, discordante o indiferente con la proposición 'la región *i* es una instancia (ejemplo) de la clase *c*'. La indiferencia surge cuando hay un intervalo de valores dentro del cual el atributo puede tanto apoyar como negar la proposición, o cuando el dato falta o es inconmensurable con la clase en cuestión (algunos atributos pueden ser aplicables a unas clases y no a otras).

Con base en esta división, se calcula un índice de verosimilitud para cada clase, de forma que para cada región se van descartando las clases menos verosímiles. El proceso se realiza en varias iteraciones, ya que el estado de regiones vecinas puede afectar el índice de las clases de la región en cuestión. El procedimiento termina cuando la mayoría de las regiones presenta una clase mucho más verosímil que las demás que además se mantiene estable. Las regiones que no han alcanzado este estado se marcarían para su inspección bien por un operario en pantalla o en el terreno si el anterior no puede resolver la duda, y tras la comprobación, las nuevas teselas del mapa actualizado se definirían como conjuntos máximos de regiones conectadas de la misma clase.

ABSTRACT

In the last decades, there has been a trend towards automated landcover mapping. Nevertheless, the overall accuracy of the maps produced in this way is normally below the user's requirements. Consequently most landcover mapping projects still rely to some extent on photointerpretation, which is less cost-effective and more subjective than the former. As the focus is increasingly shifting to monitoring rather than simple map making, there is a need to improve quantification and modelling. But this need cannot be fulfilled with traditional automated methods. The aim of this thesis is to contribute to an ongoing effort aimed at solving this dilemma by integrating into the analysis the principles of object-oriented modelling.

The traditional approach to the analysis of satellite images for landcover mapping, which is mostly based on the classification of signatures (multivariate samples drawn from pixels), leads to unsatisfactory results mainly because it hardly exploits the spatial structure of the images. In addition, this approach conceives the territory as made of distinct homogeneous surface materials, and hence cannot account for the pervading hierarchical heterogeneity of landscapes. Furthermore, the approach is in many respects incompatible with the hierarchical patch model underlying modern Landscape Ecology. The latter conceives classes as referring to types of geographic objects (patches) of a particular scale. These objects are complex compounds of objects attached to the Earth –like trees and buildings, and can be viewed as nested integrated wholes that differ somewhat from their surroundings. In being complex and nested, they are inherently heterogeneous, hence they can hardly be analysed as homogeneous materials at a single scale of observation.

The alternative approach that will be followed here is to derive, based upon the spatial structure of the image, a fine partition of the scene, in which each region exceeds the minimum size required for the geographic objects nested at the hierarchic level of interest. This partition constitutes the baseline for an object-oriented classification in which those regions are the basic units. The conceptual framework underlying such partition, and a general automated method for achieving it, are also given. The former is formalized into a three-tiered model of the territory, in which the last tier is the landcover map itself. The latter is based upon morphological segmentation and uses three consecutive techniques, namely adaptive diffusion filtering, gradient watersheds and region merging.

CHAPTER 1

Introduction

"*A map is not the territory it represents, but if correct, it has a similar structure to the territory, which accounts for its usefulness*"
Alford Korzybski, *Science and Sanity* (1933)

1.1. Introduction

The UN Conference on the Human Environment, held at Stockholm in 1972, was a first milestone in the process of raising public awareness on environmental issues, a process that reached its maturity after the Rio Conference in 1992. As a result, most governments around the world now have ministries and regional agencies devoted to environmental protection. To pursue this goal, these institutions need reliable information on relevant themes of the territory under their control. Consequently, the demand for geo-referenced environmental information has grown considerably in the last decades, at a pace akin to the development of the technologies that enable it.

Based on this information, the authorities make decisions, and the public make up an opinion that in turn influences those decisions. Thus comprehensive up-to-date geographic information on the environment is a must for a wealthy society concerned with environmental quality. This need is even more conspicuous in countries like Spain, where every year many infrastructures are built, there are land use changes from rural to urban and from agriculture to forest, and thousands of square kilometres are burnt by wild fires.

From all the layers of information involved, landcover[1] is perhaps the most important, since actions and regulations in a region are prescribed according to its nature and condition. Landcover influences markedly the productivity, vulnerability and biodiversity of ecosystems, and has a crucial impact on biogeochemical cycles, albedo, and ultimately global climate. Thus information on landcover is essential for a proper management, monitoring and preservation of our environment. This information has been traditionally presented in the form of hardcopy maps, and more recently, as digital layers integrated in a geographic information system (GIS). Within GIS, not only the information can be visualised at any scale, but combined with other themes and linked to databases for quantitative analysis.

[1] Generally speaking, landcover is the biophysical cover of the Earth solid surface. Landcover mapping units refer to both natural (and seminatural) vegetation types as well as other landcover types as e.g. agricultural fields, forest plantations and, of course, non-vegetated types like urban areas and lakes (Millington & Alexander 2000).

In the last decades, as a result of progress in Computing Science, Remote Sensing (RS) and GIS, there has been a trend towards automated mapping. Nevertheless, the overall accuracy of the maps produced in this way is normally below the user's requirements. Consequently most landcover mapping projects still rely to some extent on photointerpretation. But this procedure is not free of problems. Apart from being less cost-effective, the manual drawing of polygons involves a great deal of subjectiveness. Therefore conclusions about changes in landcover between consecutive updates made in this way are basically unreliable. As the focus is increasingly shifting to monitoring rather than simple map making, there is a need to improve quantification, modelling, and ultimately prediction (Green & Hartley 2000c). But once again, we are confronted to the low accuracy of current methods.

The aim of this thesis is to contribute to an ongoing effort aimed at solving this dilemma by integrating into the analysis the principles of object-oriented modelling. Object-orientation is the use of objects and classes in analysis, design, and programming. In particular, Object Oriented Analysis (OOA) seeks to model the world by identifying the classes and objects that form the vocabulary of the problem domain (Booch 1991). The object-oriented approach is especially useful for representing and interpreting the enduring structures of domains, integrating relevant physical entities into a coherent relational framework. This thesis deals with the linkage of three domains of decreasing complexity: landscape, RS images and landcover maps. Hence it crosses three disciplines: landscape ecology, RS digital image analysis and geographic information science. The goal is to construct an object-oriented path from patches to polygons by means of image segments, for which the first part is paved with an automated method.

The traditional approach to the analysis of satellite images for landcover mapping, which is mostly based on the classification of signatures (multivariate samples drawn from pixels), leads to unsatisfactory results mainly because it does not make use of the spatial information embedded in the images. In addition, this approach conceives the territory as made of distinct homogeneous surface materials, and hence cannot account for the pervading hierarchical heterogeneity of landscapes. Furthermore, the approach is in many respects incompatible with the hierarchical patch model underlying modern landscape ecology. The latter conceives classes as referring to types of geographic objects (patches) of a particular scale. These objects are complex compounds of sessile objects like trees and buildings, and can be viewed as nested integrated wholes different from their surroundings in one way or another. In being complex, they are inherently heterogeneous. In contrast, for the traditional approach, any arbitrarily delimited part of a classified region is still a referent of the concept conveyed by the class, provided it is large enough as to include the elements producing a typical signature of that class.

The alternative approach that will be followed here is to derive, based upon the spatial structure of the image, a fine partition of the scene, in which each region exceeds the minimum size required for the geographic objects of interest. This partition constitutes the baseline for an object-oriented classification in which those regions are the basic units. The term *object* is used here to refer to a distinct region in the image that corresponds to a patch on the ground, and that can be viewed as an instance of some *class* (i.e. as a referent of a geographic concept like e.g. 'pine forest'). The attributes taken into account in the classification include many spatial (e.g. size, shape and location of the regions) and contextual (relations between neighbouring objects, subobjects and superobjects) aspects that are not considered in the traditional approach. The conceptual framework underlying such partition, and a general automated method for achieving it, are also given.

1.2. Thesis overview

The remaining of this chapter deals with the implications of mapping landcover with satellite imagery. First, it is concisely explained what landcover maps are, how they are made, and what is the view of reality underlying such maps. Then the importance of scale is stressed and linked to the hierarchical structure of the landscape. Later on, the role of earth observing satellites in landcover mapping is set forth, and the information provided by the images they acquire is thoroughly discussed, including the influence in it of (not only spatial) resolution. This part is concluded with a detailed account about information on floristic composition.

In the last part of chapter 1, the analysis of satellite images for mapping purposes is addressed. It is suggested that the existing methods can be reduced to two main approaches, according to the order in which the identification of objects takes place: object-based (where the boundaries of the objects –patches- are defined prior to the determination of their content) and pixel-based (where the objects are retrieved by connecting after classification pixels equally labelled). The treatment of the first approach, except for a short comment on photointerpretation, is left for subsequent chapters, whereas the second is briefly described and then criticised in depth. Next, the path already followed towards object orientation is briefly studied. At last, the main points are summarized through an historic account on how we came to use the data this way, and on why this way is not longer suitable. Finally, the objectives and main contributions of the thesis are listed.

Chapter 2 is devoted to the construction of a conceptual framework for grounding object-oriented analysis of satellite images for landcover mapping. A short introduction gives way to a discussion on cognitive categorisation, in particular, the concept of partonomies (part-whole hierarchies) is introduced and linked to taxonomies through the theory of granular partitions (Smith & Brogaard 2000b). Granular partitions are ways of structuring reality (by dividing it up into meaningful chunks) in order to make it more easily graspable. Landcover maps involve the construction of two reciprocally dependent granular partitions: one over the attribute domain and the other over the territory.

Later on, the geographic objects created by the mapping activity are precisely defined and their ontological status explored through the study of the boundaries that enclose them. The vagueness inherent to these boundaries is tackled with the aid of supervaluationist precisifications on their possible location, that are condensed into a probabilistic epsilon band that in turn is approximated with the aid of a raster partition. In addition, the notions of geographic field (a regionalized variable) and measurement disk (the areal template over which class attributes are measured) are introduced.

In successive sections, the practical drawing of boundaries is addressed with an example on a hypothetical forest map. Throughout the example additional issues are discussed like the admissible size of disks, the dependence of spatial homogeneity on the observer, the existence of gaps inside objects, the need for a minimum mapping unit (MMU) and for mosaic (semantically heterogeneous) objects, and the unavoidable appearance of the modifiable areal unit problem (MAUP). This part is ended by introducing some fundamental concepts of geographic modelling, and most importantly, the topological theory of attractors (Thom 1975) is presented as the general tool with which to demarcate physical objects, in particular the minimal objects compounding an image.

Having set forth a comprehensive set of basic concepts for geographic modelling, an existing multi-tiered framework for spatio-temporal databases (Frank 2002) is adapted to construct a general landscape model for landcover mapping. The model attempts to i) justify the validity of the use of RS images to produce spatial information on landcover, and ii) formalize the conceptual foundations previously stated into a model. The model provides a framework that states explicitly how the objects created by the analysis relate to the underlying real world. It constitutes the basis of a classification method that is oriented to the construction of geographic objects from its very initial steps. Since the properties measured by remote sensors relate only indirectly to the biophysical properties used to classify landcover, the model is presented in two versions, the *idealistic* or *I-model*, and the *realistic* or *R-model*.

In the *I-model* there are unlimited resources for the measurement of landcover properties, while in the *R-model* the resources are finite. In both versions tier-0 is the commonsensical reality itself, which become measured in the next –field- tier. From the measurements, a set of continuous (I-model) or discrete (R-model) fields are created. The former represent biophysical landcover attributes, whereas the latter are constituted by satellite images. In the case of the I-model, the next and last tier (classified objects) can be directly derived from the previous one by either attribute thresholding or morphological segmentation followed by aggregation into mappable zones. In contrast, in the R-model there are severe limitations, since it is unfeasible (due to economical reasons) to measure extensively the biophysical properties intended for classification. In this case, the construction of the last tier is not trivial, since the relationship between the images and the actual values of the properties of interest is uncertain. The hypothesis linking the R-model to the I-model and the former to reality are set forth and thoroughly discussed, and a general method to obtain the last tier of the R-model is proposed. Finally, the inadequacy of the traditional approach to image classification is again addressed in the light of the new concepts introduced in this chapter.

Chapter 3 introduces an automated method that creates the starting scenario, or baseline partition, for object-oriented classification in the framework of the R-model. The method transforms the input RS image (either single or multiband) into a vector layer consisting of basic mappable zones, termed *granules*. The method is tentatively implemented as follows. The first step consists in applying a new non-linear diffusion filter that leads to a piecewise constant image where textural features are suppressed. A gradient magnitude image is subsequently computed from the filtered image. This image is then used as the input to the watershed algorithm, whose output is a primal sketch of the image with the contour of blobs (small homogeneous regions, darker or brighter –or with a different hue- than their surroundings). These blobs constitute the initial regions of a novel region-merging algorithm that aggregates them until all the regions in the partition exceed the MMU size. The dissimilarity measure used to compare regions takes simultaneously into account both luminance and chromatic contrasts. Finally, the resulting partition is vectorized, and optionally some granules attributes may be compiled in the associated database. The chapter is completed with some examples of the results achieved in real images, which are graphically displayed and discussed.

Finally, in chapter 4, the main conclusions of this thesis are summarised, and future work and open research questions are listed.

1.3. Landcover mapping: from paper to GIS

Landcover information is obtained through an inventory via which data is collected and recorded. For many years, the spatial database resulting from the inventory was a drawing on a piece of paper, the *landcover map*. Using a topographic map as background, various symbols, colours and text codes together with legends were used to display *polygons* representing semantically homogeneous patches of each landcover type that occur in the geographical area of the map. Other additional information was given in accompanying narratives. The scale was dependent on the scope of the map and the cartographic base available, typically ranging form 1:25,000 to 1:250,000.

In the last few decades, these landcover polygons were delineated manually by interpretation of aerial photography (*photointerpretation*) and then portrayed on a map using standard cartographic methods. The interpretation process consists in the identification of homogeneous regions in the image, and it is based in the visual differences that different landcover types create. The major features used for this task are texture, tonal contrast or colour, pattern, size, shape, and context. Taken together, these features make up a diagnosis that allows landcover to be mapped without having to visit every polygon on the ground. Nevertheless, there is always a need for field verification in order to assess the *map accuracy*, i.e. the frequency of labelling and/or delineation errors in the map.

The overall process of landcover mapping in the conventional way described above is too costly and time consuming as to allow timely updating. It requires a considerable amount of specialized manpower and possibly an *ad hoc* photogrammetric flight. The interval between consecutive updates for this kind of maps is typically greater than ten years, which is clearly insufficient for most applications, e.g. monitoring deforestation or urban development.

However, technical advances in the last three decades have completely changed the picture. First, computing technology has lowered dramatically the cost of data storage and increased amazingly the processing speed. Second, telecommunications now enable a massive worldwide instantaneous flux of information. An third, Remote Sensing technology has increased steadily the spatial, spectral and temporal resolution of data acquired from satellites.

Nowadays landcover maps are no longer visual spatial databases printed in a paper sheet. They are integrated in computer-based GIS, which manage large geographic databases structured in thematic layers that can be combined easily to match specific user needs. GIS enable the interaction between statistical analysis and mapping. The expanding GIS software market has also made easier and cheaper to access and use land inventories. Accordingly the demand on this kind of information is growing exponentially.

1.4. Landscape models, landcover maps and the reality behind

Landcover variation within a *landscape* (a portion of solid Earth surface on the order of 1 to 1000s of km^2) is driven by natural and/or anthropic processes, and can be modelled as continuous (gradients) or discrete (mosaics). Thematic maps represent surface variation in a highly generalised, selective and abstracted form (Ehlers, Edwars, & Bedard 1989), i.e. they describe a model of the territory rather than reality. The landscape model underlying most landcover maps is a discrete one, the *piecewise homogeneous model*. In this model, the landscape consists of a mosaic of contiguous geographic objects of irregular shape and size that cover the plane exhaustively. Each object is considered internally homogeneous at the level of abstraction of the map legend, i.e. the corresponding real surface on Earth has only one meaning in the language of the map, and this meaning is not the same than the one of adjacent objects. The homogeneity is more conceptual than real (this issue is discussed further in the next chapter). Each object is represented in the map by a *polygon*, or basic mapping unit.

These units are identified according to the similarity of a set of attributes or properties, as floristic composition, physiognomy (plant cover and size distribution), physiographic characteristics (type of soil, slope, aspect), disturbance history, etc. The number of i) attributes taken into account, ii) categories within each attribute, and iii) possible combinations of categories between attributes, constitutes the *classification scheme* of the map, i.e. the phrase book (available polygon labels) and the grammar (rules to form labels) of its language. It is selected according to the focus (e.g. land management or planning, forestry, hydrology, cadastre, biodiversity conservation, etc) and scope (local, regional or national) of the map.

Classification schemes can usually be nested in a hierarchy of increasing level of abstraction, from communities to biomes. At the same time, the landscape can be perceived as a spatially nested patch hierarchy, where patches at each level may consist of smaller patches (Wu & Loucks 1995). Such view constitutes a new paradigm in Landscape Ecology (the *Hierarchical Patch Dynamics* of Wu and Loucks (Wu & Loucks 1995)), that goes a step beyond with respect to the piecewise homogeneous model by acknowledging that pattern, process and scale are inseparable. The *hierarchical patch model* underlying such paradigm is the one that will be followed in this thesis (see appendix 1 for more information on the implications of the hierarchical structure of the landscape).

Note that the plain term ***patch*** refers to a spatial unit differing from its surroundings in nature or appearance (Wiens 1976), with no implications to its scale. That is to say that generally speaking, a patch can be from an isolated tree to an island continent. Within the patch hierarchy, one has to choose a level beneath which finer patchiness can be neglected. In the context of this thesis, that ***basic level*** consists of patches defined as **a contiguous area of similar dominant species composition and percentage of cover, occurring in an area of similar physiographic characteristics (soil, slope, aspect)**. Note that this definition implies that the surroundings of the patch show a different physiognomy and/or physiography, hence it conforms to the above general definition of patch. This unit represents the lowest level of abstraction considered (see 1.9 for a justification), and to avoid confusion will be hereafter referred to as *landcover patch*, to distinguish it from polygons, which in most maps will consist of an aggregation of these basic units.

Note also that **there exists a direct relationship between the level of abstraction of the map, its degree of spatial generalisation, and the mean size of the polygons**, since hierarchical partitions of the attribute domain create potentially hierarchical partitions of the spatial domain (Bittner & Smith 2001a). This can be better explained as follows. The world can be conceived as organised in hierarchies, so that components at different levels differ in size roughly by orders of magnitude (Salthe 1985). Since each level of abstraction is focused on a particular level of the physical hierarchy, the more we generalise the bigger the objects of interest. Understanding the what, the why and the how of the hierarchical organisation of the biosphere facilitates realizing the implications of scale in landcover mapping. In order to help myself in this task, I wrote a note on the issue that can be found in appendix 1. Although I think it is worth reading it now, those not interested in such questions can safely skip it, albeit some concepts introduced in it will be used later in the thesis.

1.5. The issue of scale

An important consequence of the pervading hierarchical structure found in nature is that knowledge on the real world is also layered. To put an example, a plant pathologist does not need to know the configuration of molecules inside the cell of an abnormal tissue to diagnose a disease, and neither need a forester plant physiology to estimate the height growth of a tree, and neither a meteorologist forestry to forecast regional weather. Since knowledge on the real world is acquired through observation, the consequence is that each layer of knowledge has its own appropriate *scale* of observation. The latter is normally done through an instrument, as a microscope or a camera.

The range of scales at which the observation is useful depends on the size of the *objects of interest* of each particular layer of knowledge. Given a particular object like e.g. a tree, the upper bound of this range is the largest scale at which the object is still recognizable, while the lower bound is the minimum size of the field of view that completely encompasses the object. When there is no *a priori* information about the size of the objects, or when the variability in size of the objects is too high, it is not trivial to establish the appropriate scale of observation (Lindeberg 1994).

This concept is closely related to John Wiens' (1989a) *domain of scale*: a region of the scale spectrum within which patterns and their relationships with underlying processes either do not change or change monotonically with changes in scale. Each domain corresponds to the range of scales at which the objects nested in a particular level of the hierarchy are *visible*. The transition, or *scale threshold*, from one domain (e.g. the set of scales at which individual trees are still observable) to an adjacent domain (e.g. where the trees are no longer visible and the forest become apparent as an object of its own), may be relatively abrupt and characterized by complex non-linearities arising from dissolvence/emergence processes (appendix 1), much like phase transitions in physical systems. Thus relationships between variables may not be easily predictable between domains, and this may cause erroneous inferences like the *individualistic fallacy* (to derive properties of the aggregates from the ones of the individual components) or the *ecological fallacy* (the other way round) (Alker 1969).

It is important to note that scale thresholds do not occur simultaneously for all the objects compounding a given level of the landscape hierarchy. This is due to the fact that these objects are generally sufficiently variable in size that the different levels in the hierarchy overlap in size (Woodcock & Harvard V.J 1992). This view is supported by the relatively gentle slope of the scale variance graphs that Townshend and Justice (Townshend & Justice 1988) found when they tried to show that there is still information at virtually any spatial resolution as imagery is degraded to coarser scales. The consequence is that there is no single optimal scale to study landscape at a given level of abstraction.

The scale range of observation depends on the instrument and is bounded on two sides: the *inner scale* is the smallest detail seen by the smallest aperture (a CCD element in SPOT HRVIR or a cone or rod in the human eye); the *outer scale* is the coarsest detail that can be discriminated, i.e. the whole image, scene or field of view (Haar Romeny 1997). The outer scale can be seen as well as the geographic extent of the study area, since several images can be mosaicked to cover the whole area if the latter is larger than a single image. The inner scale is equivalent to *spatial resolution* in Remote Sensing literature, and is closely related to the *ground-projected instantaneous field of view* (GIFOV) of the sensor. The *pixel size*, i.e. the *ground sampling interval* (GSI) or ground area corresponding to each data element in Remote Sensing standard products, is chosen to match the GIFOV, although the *point spread function* (PSF, related to the variable sensitivity of the detector within the area of integration) of the sensor usually yields an *effective* GIFOV coarser than the geometric one (Schowengerdt 1997). Nevertheless, **within this thesis, pixel size has the same meaning than spatial resolution**, although this does not necessarily mean that it coincides with the size of the smallest ground object that can be reliably detected (Davis & Simonett 1991).

The classic cartographic scale, i.e. the constant ratio between distance between pair of points on the map and distance between the same pair of points on the Earth surface (Goodchild & Quatrocci 1997), is chosen so that the objects of interest can be clearly seen at glance on the map, and the smallest size of them (the *minimum mapping unit*, MMU) still seen by the naked eye. Thus the extent and level of detail of a map are somehow related to respectively the field of view and maximum resolution of the human eye, so that the density of visual information portrayed in maps of different scale is roughly the same (Frank & Timpf 1994). To end up this account by connecting it to the previous section, it can be said that the scale of the map is determined by the layer(s) of knowledge about which the map is aimed to inform. Each scale corresponds to a certain level of abstraction of the Earth surface. Therefore each level of abstraction has a direct correspondence to the size of objects of interest, landcover patches in our case. Now we turn to the modern means that provide information about them.

1.6. The role of Earth Observation Satellites in landcover mapping

Remote Sensing[1] (RS), as understood in this thesis, is the measurement of some electromagnetic properties of the Earth's surface (land and oceans) through sensors mounted on aircraft or satellites. Satellite Remote Sensing is generically known as Earth Observation (EO), and the measurements supplied –usually in the format of digital images-, EO data. Land EO began with the launch by NASA of the ERTS-1 (Earth Resources Technology Satellite, later renamed Landsat-1) in 1972. The Multi Spectral Scanner (MSS, four bands) on board this satellite provided for the first time a consistent set of synoptic (185 km swath[2]), high resolution (80 m) images to the scientific community. Since then, the number and capabilities of EO satellites have grown steadily. Currently there are more than 30 civilian satellites providing data of 30 meters or better resolution (Stoney 1997).

[1] The term "remote sensing" was coined by Dr. Evelyn Pruitt in the 1940's when she was with the U.S. Office of Naval Research. The term generally implies that the sensor is placed at some considerable distance from the sensed target, in contrast to close-in measurements (Short 2002).
[2] The swath is the width of the strip of the Earth's surface imaged in each orbit.

The present EO satellites provide an alternative source of information on landcover more cost-effective than aerial photography. First, EO data are cheaper. For example, it takes some 750 1:50,000 aerial photos to cover a 185x185 km² Landsat 7 scene that costs only 700 euro (2 cents per km²) and has a slightly inferior level of ground detail (15m resolution in panchromatic band). Even 1m resolution satellites as Ikonos costing 20 USD/ km² are still cheaper if the task requires a new flight.

Another benefit of EO data is their digital format. Spatially the data form up a matrix or *raster*[1] composed of cells, each one having a *digital number* (DN) which may be related, normally after some correction or calibration, to the biogeophysical characteristics of the piece of land to which that cell corresponds. The result of displaying (or printing) this data is a grey-level image or *band*. This is done by relating each raster cell to a picture element (*pixel*) of a screen through a *Look Up Table* (LUT), which assigns a discrete brightness level to each DN. This association is the reason why the term *pixel* is equated to *data element* in EO data. The LUT entries are usually chosen as to enhance the image contrast. Since colour monitors have three guns (Red, Green and Blue), it is possible to display three bands at a time, forming a RGB colour composite (Pinilla 1995).

Multispectral images typically consist of several bands, each one corresponding to a particular interval of the electromagnetic spectrum to which the respective detector is sensitive. The location of these intervals in the spectrum is selected as to produce a good discrimination between different ground features. This digital nature enables a number of quantitative computer-assisted analysis that yield objective estimates about landcover and other features of the Earth surface that can subsequently be integrated into a GIS.

Finally, another important advantage of EO data is the readiness with which they can be purchased. There are several private companies that distribute in a commercial basis data from satellites operated by themselves or by space agencies, and thus relieve users of all data acquisition problems. Most of these satellites fly almost along meridians in a sun synchronous *Low Earth Orbit* (LEO, ranging from 250 to 800 km altitude), permitting repeat intervals from 3 to 60 days depending on the swath of the image. This means that for most places of the

[1] Altough strictly speaking, the term 'raster' refers to a scanning pattern of parallel lines that form the display of an image projected on a cathode-ray tube of a screen, it is widely used instead of grid (a pattern of regularly spaced horizontal and vertical lines forming squares on a map, a chart, an aerial photograph, or an optical device, used as a reference for locating points).

world, it is possible to purchase from 1 to more than 20 cloud-free optic high resolution images a year, depending on the cloudiness at that site. The latter constraint does not hold for Synthetic Aperture Radar (SAR) images, although the information on landcover that can be brought forward through their analysis is more limited.

In short, the multitemporal and multispectral capabilities of EO satellites and the inexpensiveness of the data they provide make them superior to conventional aerial photography as a source of up-to-date landcover information, even more when there are already satellites providing resolutions equivalent to the detail of the photos. The problem now is how to derive this information.

1.7. The information content on landcover of EO data

Information comes ultimately from physical order. Order implies differences, that is, non-uniform distributions of matter/energy. Certain systems having sensors can take advantage of these differences for cognitive purposes, provided they can detect them. From all the set of detectable differences (**latent information**), only the relevant ones are used by those systems to construct a handy representation (**structural information**) of the sensed scene, that has to be formalised into a model in order to be communicated. If the system has cybernetic –communication and control- capabilities, and the model is taken into account in its decision-making process, then it becomes **functional information**. The meaning of that information is the prediction of the success of the selected action, and its value depends on the consequences of that action. If the result of the action is the one expected, then this information is stored as (reusable) **knowledge**.

These are the conclusions I have drawn from a limited inquiry I have made on the concept of information. I felt it was apposite to elaborate on it, given the lack of a clear definition in the literature. Interested readers can find the full text in appendix 2. Another conclusion is that satellite data have no fixed information content, it rather depends on the task at hand. In any case, it is the structure of the image, formed by luminance and chromatic changes taking place across it, the only bearer of (latent) information. From all (either spatial or spectral) patterns that can be observed in an image, only those coming from objects that are to be detected and modelled are relevant. Thus the goal of image analysis is to make explicit these patterns by drawing (either spatial or spectral) boundaries separating the objects of interest. Later it is assumed that the objects created by the analysis have a fixed correspondence (given a fixed set of viewing, illumination and atmospheric conditions) with the objects in the real world that gave raise to the (spatial and spectral) structure of the image.

Before going on, I would like to make clear the concept of the **spectral structure** of a multi-band image, since it is not as intuitive as the spatial one. Such structure is derived from *signatures* usually drawn from individual pixels. A signature is an *n*-component vector where each component or dimension corresponds to a band, and the value at each component is the DN of the pixel in the corresponding band. Such signature can be plotted as a dotted curve (that obviously can also be interpolated to draw a continuous curve) where each dot is the value in one of the consecutive spectral bands of the image (see figure 1-3 in page 21). Since this is the form usually adopted by the measurements of desktop spectrometers, hence the name of signature.

This vector can be also viewed as the coordinates of a data point in an *n*-dimensional stochastic space[1] usually referred to in the Remote Sensing literature as the *feature space*, which here will be termed the ***data space***. The non-uniform arrangement of signatures (which usually tend to cluster into more or less discontinuous regions) within this space forms a structure that is seized in the analysis. Albeit the adjective *spectral* is more restrictive than *feature*, the latter is more ambiguous (it is also used for ground objects), therefore the former will be preferred in this thesis when talking about the structure of the data space. This does not mean that the discourse on multi-band images holds only for optic ones; rather, it could be applied to any multi-band image, no matter the nature of the measurement. Finally, note that instead of individual pixels, signatures can also be drawn from groups of connected pixels. For this purpose, the image is previously partitioned into homogeneous regions and the mean value of the pixels inside each region is used as a signature. In this case, the data space is less densely populated than when each signature consists of a single pixel.

Last but not least, it is important to note that, following the discussion in appendix 2, **information is not extracted but produced during the analysis**. The result of the analysis is the imposition of a simpler, meaningful structure upon the intricate original one, that is, a formal representation (a model) of the structure of the image. This model is dependent not

[1] Where the location of the data point can be interpreted as the result of an stochastic process. A stochastic process is an *a priori* indeterminate process that is characterized by a distribution. If there are many and independent elemental effects, the distribution of the process is a Gaussian (Koch 1988).

only on the data alone, but also on the definition of the type of objects we want to foreground and on how detailed the representation should be. Different choices during the process of analysis will yield different representations (and hence different information) from the same data set.

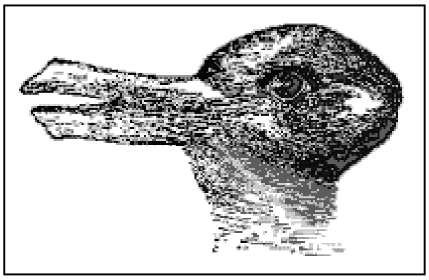


Figure 1-1. Duck or hare?

In order to emphasize the former assertion, note that the etymological origin of *informare* ('to give form', in Latin) implies the realisation of structure upon some material. Thus the image can be seen as the raw material upon which the model is carved out. The final form of the representation is dependent not only on the material but also on the analyst. However this creative freedom is by no means unrestricted. In the example of figure 1-1, if the task is to identify an animal you could argue whether it is a duck or a hare, but if you see a cat, you would probably be invited to visit a psychiatrist. Hence structural information is subjective, but it has also a structure similar to the one of the piece of reality it refers to. The latter property is the hardest constraint to the output of a sound image analysis, and indeed is this isomorphism[1] what accounts for the usefulness of the result (recall the opening quotation of this chapter).

So now, what is the information content of EO data on landcover? Assuming the piecewise model of 1.4, and having set a certain level of generalisation according to our objectives, the objects of interest for the study of landcover are a set of geographic objects to be represented in the map by polygons. The goal is to identify these objects within the area of the map and get answers about the size, shape, location and inherent properties of each one of them. The properties to be studied are those considered in the classification scheme (vegetation physiognomy, floristic composition, and so on). The information content of a data set can only be defined in relation to this scheme. Intuitively it would be the fraction of questions identifying the objects, as defined by the scheme, that can be answered reliably through the analysis of these data. Since there is a surrogate relationship between the defining properties of landcover and the ones measured by remote sensors, the analyst may answer some of these questions using EO data. Although the mutual implications of this relationship will be

[1] Strictly, a term in mathematics for an exact correspondence between both the elements of two sets and the relations defined by operations on these elements. Obviously here the correspondence refers only to the elements of the territory that are represented in the image, otherwise we would be saying that RS images are territories!

addressed with further detail in the next chapter, some considerations on its nature must be done now.

Satellite sensors measure Earth surface electromagnetic properties, such as the reflectance (indirectly, see footnote on next paragraph) of solar radiation in different intervals of the spectrum, in optic sensors, or the fraction of a power pulse emitted by a SAR instrument that is scattered by the surface back to the radar antenna (backscatter), in active microwave sensors[1]. The analysis of these measurements can bring forward information on our objects of interest, since the electromagnetic behaviour of a landcover patch is dependent on a certain number of characteristics that can be linked directly, through allometric equations[2] or by other means to the properties considered in the map, as we shall see in the next paragraphs.

In the case of optic sensors, canopy apparent radiance[3] as observed from a satellite is a function of: i) structural properties such as leaf area index (LAI, half the total green leaf area in the plant canopy per surface unit), leaf size, shape and inclination, horizontal and vertical distribution; ii) optical properties (reflection, absorption and transmission from, within and through tissues, which depend on their biochemical composition and thickness) of leaves, other vegetation elements (bark and flowers) and soil; iii) incidence and viewing angle; and iv) atmospheric conditions (Asrar 1989).

In the case of SAR instruments, radar backscattering from vegetated areas depends on three groups of factors: i) target structural attributes (soil roughness and canopy architecture, that is, size, shape, distribution and orientation of scatterers: leaves, branches and trunks); ii) target dielectric properties, mainly controlled by the moisture content of soil and canopy; and iii) wavelength, polarisation and look angle used by the system (Ulaby & Dobson 1989). The wavelength λ of the emitted pulse determines the penetration depth of the waves in the imaged target. Thus the scattterers for X-band (λ=3cm) will consist of the first leaves on top

---

1 Readers not acquainted with radar imagery and willing to get some basic insight may refer to "Principles of radar imagery", FAO Remote Sensing series No 62, 1989.

2 Equations commonly used in Biometry, in which one part or characteristic of a living entity (e.g. the Leaf Area Index (LAI) of a forest) can be expressed as a power function of another part or characteristic (the Basal Area (BA) of that forest).

3 The apparent radiance is the amount of power density coming from the sun scattered by both the canopy and the atmosphere (path radiance) in the direction of the sensor. Radiance has units of Watts per square meter per steradian, whereas *reflectance*, an inherent property of surface materials describing what proportion of the incident energy is reflected, is dimensionless (Richards 1993). In principle, it is possible to derive estimations of reflectance from radiance measurements. NASA is currently testing a new instrument on board EO-1, the Atmospheric Corrector (AC), that will provide accurate correction for atmospheric variability (primarily water vapor and aerosols), enabling accurate reflectance measurements for land imaging missions.

of trees, whereas for L-band (λ=23 cm) they will be thick branches and trunks. Apart from this, volume scattering tends to depolarize the pulse, so that polarization may provide information on the shape and orientation of the scatterers (FAO 1989). These features will be exploited by the European TerraSAR mission (http://www.infoterra-global.com/terrasar.html), to be launched in 2005. TerraSAR will make near-simultaneous observations of the Earth in high spatial resolution up to 1 m in X-band and with full polarimetry and high radiometric resolution in L-band.

Apart from backscatter intensity images, *coherence* images, derived from pairs of SAR complex (phase and magnitude) images acquired over the same scene at different times (therefore from slightly different orbits), deserve a comment. Coherence is a statistic for describing the quality of interferogrammes (that record the differences in phase between the two images), which may be used for landcover mapping complementarily to backscattering intensity. It can be interpreted as the fraction of power scattered by unchanged parts of the scene. Coherence over a certain area is determined by phase decorrelation within an averaging window, often caused by random displacements (due to wind) of the contributing scatterers, or by a random thickness change of an intervening dielectric medium (due to e.g. rain or snow). Therefore information in coherence images comes from the differential behaviour of each landcover type, in terms of relative permitivity and geometry of the scatterers, when exposed to different wind and moisture conditions (Hobbs, Ang, & Seynat 1998).

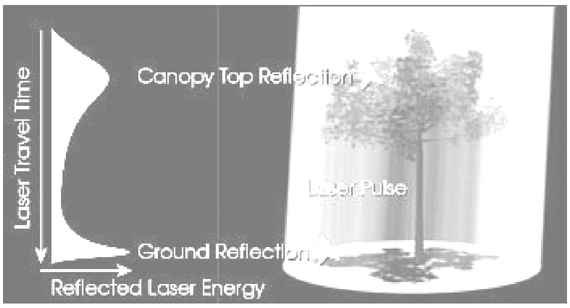


Figure 1-2. Lidar waveform diagram (Image by Robert Simmon, NASA Earth Observatory)

Finally, it is important to mention that a spaceborne LIDAR (Light Detection and Ranging, that is, an improved laser altimeter), will be available shortly. The Vegetation Canopy LIDAR Project (VCL), the first NASA Earth Systems Pathfinder Mission (http://essp.gsfc.nasa.gov/vcl/), due to launch on 2000 but delayed because of financial problems, is designed to provide a global database of forest vertical structure (layering) and tree height. Lidar sensors measure elevation by bouncing laser light off of a surface and measuring the time the light pulse takes to return. By also recording the shape of the "waveform" of the returned signal, the location —in the space between the ground and the tree tops- where the foliage, trunk, and branches are concentrated can also be estimated (figure1-2).

Coming back to the point, the relative contribution of landcover properties to the actual remote measurement together with the stability of that contribution define the strength of the surrogate relationship. If the contribution is high and is always the same, the conclusions drawn from data regarding the property of interest are likely to be right. In the other extreme, if the contribution is low and/or changes randomly from one location or time to another, there will be a lot of uncertainty about the conclusions drawn, if there are any at all.

For most instances, the situation is an intermediate one. A good example would be the normalised difference between the near infrared band and the red band of an optic sensor over a timber forest. This parameter, known as the *Normalised Difference Vegetation Index* (NDVI), is highly correlated to LAI, which in turn can be used to estimate basal area, and the latter together with the mean tree height are the main variables used to estimate timber growing stock, that could be a key element to map the forest into stands for management purposes. However this correlation is not stable, it tends to decrease when LAI is high due to the early saturation of the red band (Peterson & Running 1989).

In addition, the more factors external to the map classification scheme are involved in the sensor response (atmospheric and illumination conditions, sensor gain, variable extrinsic factors affecting land objects, as snow or drought), the higher the uncertainty of the information produced with the data. The reliability of that information decreases with the magnitude of these external factors. An extreme example of this would be an optic image of a completely cloudy scene, in which there is no information at all about ground objects. Finally, there are also factors related to the resolution of the measurement, which will be discussed in the following section.

To sum up, the information content of an image can only be defined in relation to a specific task with definite objects of interest. Furthermore, it can only be measured after having performed the analysis, since it is in this process where the information is actually produced. Apart from the analysis itself, this *ex post* content relies on the strength of the relationship between the properties measured by the remote sensor and the ones defining the objects of interest. Hence it is dependent on the relative contribution in the sensor response of i) landcover properties taken into account in the map classification scheme; and ii) properties external to that scheme. The reliability of the information derived from the data is directly proportional to the weight of i) and inversely to the one of ii).

1.8. The impact of resolution on the information content

Not only the nature of the sensor measurement affects the information that can be derived from it, but even more markedly, the level of detail or *resolution* of that measurement will determine its potential to bring forth useful information on the Earth surface through analysis. Resolution has four dimensions: radiometric, spectral, spatial and temporal, that will be discussed in the next paragraphs.

Radiometric resolution is the number of bins into which the continuous range of possible sensor responses is quantized. The photons reaching the detector during the exposure time (in the order of milliseconds) at each sampling position (pixel) are converted to an electrical signal and then quantized to a discrete number that is expressed in binary digits (*bits*). As with all digital data, a finite number of bits (BD, the *bit depth*) is used to code the measurement, so that the DN can be any integer in the range $[0, 2^{BD} - 1]$ (Schowengerdt 1997). This *dynamic range* varies from the 8 bits [0,255] of Landsat and SPOT products to the 16 bits [0,65535] of Envisat ASAR PRI (precision radar image).

From the four aspects of resolution, the radiometric one is perhaps the less important with regards to information. Visually, it is virtually impossible to distinguish e.g. a 5 bit (32 grey level) rescaled image from the 16 bit original one. From the quantitative analysis perspective, (Narayanan, Sankaravadivelu, & Reichenbach 2000) showed that the results of classification over a TM scene did not change significantly by reducing the bit depth from 8 to 4 bits, suggesting that a great deal of memory can be saved even before applying any compression on the data set. The reason behind such an apparent waste is that the sensor must be able to measure very different situations without reaching saturation and without losing discrimination power over a relatively homogeneous scene.

Spectral resolution applies only to optical (and thermal) remote sensing. Multispectral (several bands) and hyperspectral (tens of bands) sensors split the reflected solar light into multiple paths with the aid of prisms, and have different spectral filters in each path, so that several to tens (even hundreds) of bands can be created, each one displaying the amount of energy received in a particular interval, or band, of the electromagnetic spectrum. Remote Sensing spectrometry relies in the fact that reflectance values for each material vary according to the wavelength distribution of the incident radiation. Since this variation is unique for each material, the response of the sensor in different bands can be used to identify the material from which the imaged surface is made. Therefore, the greater the number of bands and the narrower their width, the higher the spectral resolution and the more accurate the identification can be.

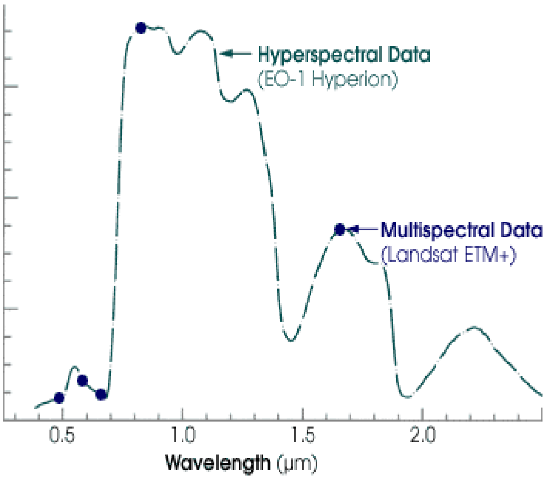


Figure 1-3. Hyperion's 220 bands (hashed line) provide a more accurate depiction than the broad bands of Landsat (dots) . (Graph by Robert Simmon, NASA Earth Observatory)

Spaceborne hyperspectral imagers are already available. The Hyperion instrument, on board NASA EO-1, provides 220 contiguous spectral bands (from 0.4 to 2.5 µm, each band of 10 nm width, see figure 1-3) with a 30-meter pixel size. From the hyperspectral data, it is possible to construct a spectral curve that i) can be matched with spectral signatures of pure materials (*endmembers*, that is, the centroids of spectral classes) available in spectral libraries; or ii) can be searched for specific reflectance peaks and absorption troughs indicating the presence of special materials; or iii) can be analysed with some *spectral unmixing* technique if the spectral curve is suspect of being a mixture of several materials.

Spatial resolution can be equated to the pixel size, i.e. the ground sampling interval (GSI, see 1.5). Geographic space variation is *regularized* through the convolution of the signal coming from individual ground objects within this sampling field (Jupp, Strahler, & Woodcock C. 1989). In other words, the measurement at each pixel is an averaging (weighted by the PSF) of the radiance incoming from the ground area corresponding to that pixel. If the objects of interest are smaller than the pixel size, it is virtually impossible to derive any conclusion about their shape and size as individual entities (e.g. the crown size of individual trees from a 30 m resolution Landsat image, or the boundaries of a crop paddock from a 1.1 km resolution AVHRR image). If pixels are in the same size range of objects, the correlation between object properties and sensor response will be maximal for pixels fully comprised within objects, but in turn there will be a high proportion of pixels including several objects with different properties where the mixture will degrade the correlation (Woodcock & Strahler 1987).

Finally, if the objects of interest exceed by far the pixel size, the parts (e.g. trees) composing the object (landcover patch) become apparent, and the measurements of individual pixels can be related more readily to the properties of the parts than to the ones of the whole, complicating the analysis of the data. In turn, precise estimates of the shape and location of the objects of interest can be drawn. The first case in the previous paragraph is referred to in the literature as the *L-resolution* model, whereas the one in this paragraph is the *H-resolution* model (Strahler, Woodcock C., & Smith 1986), where H and L stand respectively for high and low. As these authors pointed out, these concepts can be extended to time as well as space, and will be referred to as *H-frequency* and *L-frequency* models, to distinguish them from the previous models.

Temporal resolution refers to the repeat frequency of the measurement. This frequency has an impact on the information that can be derived from the data, since the electromagnetic properties of a given area are by no means stable. Change happens over time scales from seconds (leaves orientation changes due to wind, that affect e.g. *interferometric coherence*) to centuries (e.g. successional processes) (Davis & Simonett 1991). H-frequency implies that the observations are repeated at higher frequency than significant changes in the measured properties (e.g. daily NDVI from AVHRR), so that abnormal abrupt changes can be located in time (e.g. floods, fires). L-frequency means that the repeat interval of the sensor is too long to detect changes of interest whose effects disappear before the next observation (e.g. a flash flood remaining only a few days may not be detected by the 35-day repeat cycle ERS-2 SAR, and the same can be said for Landsat TM about seasonal vegetation changes in a site where cloudiness allows cloud-free images only in the dry season).

As with spatial scale, the information content on landcover of a multitemporal series of satellite images from the same scene depends on the properties about which information is searched. If the aim is e.g. to monitor processes as urban development or deforestation, a series of SAR images acquired in the same month of consecutive years is appropriate, but it cannot bring more information than a single image about landcover attributes that depend on the seasonal variability of vegetation or the year cycle of different crops. Conversely, a series of 10 SAR images of the same year may bring enough information as to discriminate between deciduous and evergreen forest or between different crops, but if used to monitor deforestation, the observed seasonal changes can led to wrong interpretations.

1.9. An example: information on floristic composition from EO data

In summary, the information content of EO data about a particular theme depends on **i)** the type of analysis performed (which will be addressed in next section); **ii)** the size and defining properties of the objects of interest of that theme, **iii)** the relative contribution of these properties to the sensor measurement, **iv)** the relative contribution of other factors not related to these properties, and **v)** the resolution (spatial, temporal, spectral and radiometric) of the measurements. The impact of these factors will be illustrated with an example about an important property of landcover patches: their floristic composition. But first a definition of *floristic composition* within this context is needed.

The characteristics of ecosystems are determined by the primary trophic level, the vegetation, so that the latter can be taken as surrogate of the whole ecosystem (Graetz 1990). For this reason landcover classifications use vegetation as the principal geographic phenomenon to classify the Earth surface into landcover types. Each landcover type is an abstraction of a collection of geographic objects (represented in the map by polygons) summarising the set of common attributes shared by these objects.

At a broad level of abstraction, vegetation can be described by its overall appearance, i.e. its physiognomy, without entering in floristic details. The most significant physiognomic features are the height of the uppermost, or *dominant*, existing stratum (that would yield e.g. forest-shrubland-herbaceous categories), and the proportion of ground covered by that stratum (e.g. dense forest-woodland-sparse woodland). A lower level of abstraction would require additional information on the phenology (deciduous v. evergreen) and shape (broad-leaves v. needles) of leaves most frequently found in the dominant stratum. If a further level of discrimination is required (e.g. we want to know the quality of timber we could extract from a conifer forest) then floristic information is needed, ranging from only identifying the dominant species (this would be enough for the former example) to a complete floristic inventory with the occurrence and frequency of plant species in all the strata (this would be required for e.g. a biodiversity study).

On the other hand, the value of spatial data decreases with the outer scale (geographical extent), and its cost increases with the inner scale (resolution). Vg the scale of observation appropriate for a comprehensive floristic study requires expensive and time-consuming field surveys. These cost/benefit considerations make unsuitable to base landcover mapping in units finer than the *landcover patch* defined in 1.4. Therefore studying *floristic composition* within this context means to identify the main species occurring in the dominant stratum of the vegetation, which can be observed by EO satellites as opposed to the understory.

The size of the diagnostic elements ranges from the few centimetres of the herbs of a grassland to the several tens of meters of the trees of a forest. The scale of observation ranges from millimetres (as for e.g. counting the stamens of a flower) to tens of meters (e.g. to appreciate the shape of a tree). The properties that lead to the identification are related to the morphology (type of flowers, fruits, leaves, bark, ramification, and crown) of parts of the object or of the whole object (plant individuals). Once identified, a set of other properties can be brought forward by previous knowledge accumulated by the sciences of Botany and Ecology, as timber quality, periods of flowering and fructification, requirements in terms of light, rain, temperature and soil, association with other species, etc. Conversely, an expert can use this knowledge to make a heuristic conjecture on the species composition when the scale of observation is inadequate. This allows e.g. a forester to identify a tree species from several hundred meters when there is enough contextual evidence.

The contribution to EO measurements of the properties used in the identification is low in comparison to other factors. There are some relationship between inherent properties of different species as the optical thickness of individual leaves and the response of an optic sensor, or as the size and shape of leaves and the radar backscatter, that can be exploited to derive floristic information. The problem is that the measured response may be overweighted by factors that are not intrinsically specific, as the amount of green matter in the case of optic sensors or the relative dielectric constant in the case of radar.

Despite this shortcoming, the focus of analysis has been on identifying the spectral signatures of the classes of interest. The reason behind this approach is that the spatial resolution of EO data has been too low[1] to identify single trees by their shape or leaves (Landgrebe 1999). As a matter of fact, the only spatial difference traditionally taken into account in quantitative analysis of EO data has been *texture*[2]. But, even acknowledging that different landcover types may yield distinct textures for a given pixel size, the problem is that it is not trivial to determine, for each location in the image, the region from which the texture measure should be derived (Lobo 1997). This issue deserves a one-paragraph detour.

Texture measures are usually produced by computing a statistic (e.g. variance) from the DNs inside a rectangular template (e.g. a 5x5 pixel window) centred at each pixel. The resulting image is subsequently included in the analysis as another 'spectral' band. Apart from being a mixed bunch, this approach has three serious problems. First, window-derived measures cannot distinguish bare texture from relevant geometrical patterns, since the technique is aspatial, i.e. it does not take into account the position of each value within the window (Ferro 1998). Second, depending on the size of patches relative to the pixel, there may be a lot of locations for which the respective window will include pixels from an adjacent patch, yielding a confusing measurement. And last but not least, as a consequence of the former, classification accuracy using texture is far more dependent on the size of the computing window than on the type of texture measure employed (Marceau et al. 1990).

Turning back to the vegetation classes commonly studied with EO data, they usually consist of broad categories (broad leaved forest, coniferous forest, grasslands, etc), but sometimes they include some floristic element in the definition (e.g. *Fagus sylvatica* forest). In any case, the underlying assumption of the spectrometric approach is that each class has a consistent signature that can be separated from the rest, that is, patches of the same class will show a similar response, and this response will not be similar to the one of patches of a different class. Unfortunately, this is seldom the case: even correcting for atmospheric and illumination effects, one cannot count on a vegetation class having a consistent spectral signature, no matter how high the spectral resolution is. The radial structure of classes (discussed in 2.2.3),

[1] It has been historically unfeasible, for both economical and technical reasons, to reach higher resolutions. Higher spatial resolution implies huge data volumes and faster downlinks, improved optics and electronics, and precise platform control. On the other hand it leads to reduced swaths, hampering a quick coverage of large areas.

[2] Texture is the local variation of brightness in an image caused by the irregular response of spatial structures that consist of recurrent elements (e.g. leaves in IKONOS or trees in Landsat) that cannot be resolved by the sensor because of their reduced size in comparison to the sensor's GIFOV. If one thinks of a digital image as a tangible surface given by DN(x,y), then texture could be assimilated to the roughness sensation of that surface when passing a finger across it (Schowengerdt 1997).

which may include in the same group very different settings, precludes the very existence of 'class signatures'.

Some of the parameters affecting vegetation reflectance relate specifically to particular species, like the morphology of individual plants, including the size, distribution, orientation and optical properties of leaves. But many others do not, like the structure of the canopy, the nature of background soil and/or understory, time of the year, growth stage (even-aged stands), plant health, and moisture content (Armitage, Weaver, & Kent 2000). Besides, solar light interaction between the different parts of the vertical profile of a vegetated area is a rather complex one, involving multiple scattering and selective transmission among the parts.

This interaction is even more intricate when the canopy is not dense enough and the trees do not cover totally the ground, as is the norm in Mediterranean landscapes. This means that the overall response cannot be reduced to the summation of the individual responses, making spectral unmixing unfeasible when the pixel size is in the same range than this interaction. The classical pattern of reflectance from vegetation (low in the visible and high in the near-infrared) depicted in textbooks is based on the response of single leaves. Therefore it can only be assimilated to the one of the canopy if LAI is high enough (Curran 1985). In this case, the identification of canopy species might be feasible with an adequate spectral resolution. Under other circumstances, the correspondence between spectral response and species composition would be difficult to identify.

Regarding temporal resolution, the different rhythms between species in response to seasonal changes (defoliation and flowering) create variations in the sensor response that can be successfully used to discriminate between species when H-frequency data are available. Hence intra annual spectral-temporal profiles, together with knowledge of crop calendars and vegetation phenology, can be used to map different landcover types and even individual species (Reed et al. 1994). However, there are three limitations to this approach. The first one is due to inter annual variations of phenological events, related to changes in accumulated precipitation and temperature from one year to another (Weber 2001) that may alter the known sequence upon which the identification is done. The second refers to variations regarding the general profile due to latitude and/or altitude. The third and last is the difficulty to get a complete yearly profile due to cloudiness, so that some key observations may be missing.

The latter impediment does not hold for radar images, although (Proisy et al. 1999), studying the foliage seasonal cycle of a mixed forest using ERS SAR (C-band), concluded that it could not be reliably detected, partly due to the strong contribution of branches in the backscatter. As a point of fact, it is widely accepted that it is not possible to identify individual tree species with the current spaceborne radar systems (Quegan et al. 2000;Wagner, Vietmeier, & Schmullius 2000), although the advent of multifrequency multipolarimetric radars may change the picture.

In short, **the information content from EO data on floristic composition**, that is, the possibility of identifying individual canopy species through their analysis, **is poor**. This is mainly due to the inadequate scale of observation and to the variable contribution of species intrinsic properties to the sensor response. **However, different combinations and abundance of species produce differing patterns in the image** (Armitage, Weaver, & Kent 2000). These differences could be better exploited if the focus of the analysis shifts from the spectral signature of classes to the more realistic one of the spectral signature of individual patches, as proposed by Smits and Annoni (1999). This shift implies that an estimation of the *form* (size, shape and location) of the patch has to be available before inquiring about its *substance* (physiognomic and floristic properties).

1.10. Analysing EO data to derive information on landcover

The goal of image analysis, as stated before in relation to the study of landcover, is to produce a handy representation of the imaged scene by identifying in it the objects of interest, landcover patches in our case. The identification of an object involves two aspects: i) its *form*, i.e. the boundaries of the object; and ii) its *substance*, i.e. the constituents from which the object is made. I consider the *form* aspect as related to the spatial structure of the image, whereas *substance*, although not alien to the former, has more to do with the spectral structure. In any case, the result of the analysis is a thematic map of the territory under study that portrays the type of land cover object that is expected to be found at each location. Again, the assumption is that the objects created by the analysis have a definite correspondence with the objects in the real world that gave raise to the (spatial and spectral) structure of the image.

Two further assumptions will be made within the scope of this thesis. The first one is that imaging geometric distortions (due to relief effects and off-nadir viewing angles) can be adequately rectified, and that referencing to a geographic planar projection system (as e.g. UTM) is performed accurately, so that **positional accuracy is good enough** (that is, that the actual coordinates of the centre of a given pixel are within that pixel). The second one is that **the final format of the map is vector** rather than raster, in other words, that the *minimum mapping unit* (MMU) is several times larger than the pixel. Under this circumstance, it is far more economical (recall the MDL principle introduced in Appendix 2) to deal with polygons than with individual raster cells. The MMU concept is the cornerstone of the conceptual framework proposed in this thesis, and will be further discussed in chapter 2.

Having taken for granted the correct pre-processing of the available images (co-referencing, resampling and geocoding), it can be said that there are **two main approaches to image analysis**, according to the order in which the identification of objects takes place: **object-based** analysis (or *first form, then substance* methods) and **pixel-based** analysis (or *form from substance* methods). In the first case, a segmentation (i.e. a partitioning into non-overlapping regions, or *segments*) of the image is carried out as a first step in the analysis. The resulting segments are subsequently classified using *inter alia* their spectral properties, which are usually estimated from the mean values of the pixels belonging to each segment. After having been classified they become objects endowed with an enduring identity. In the second case, pixels are labelled (i.e. classified) according to their location in the multidimensional data space, without considering (apart from perhaps adjacent pixels) their position within the image. Afterwards, objects are defined as sets of connected pixels with the same label. This latter step often involves a previous change in the label of some pixels as to enable the vectorization of the classified image, otherwise unfeasible due to the lack of spatial consistency of the result. This *patching-up* process is commonly referred to as *post-classification*.

Object-based analysis will be discussed with greater detail in chapter 2, but before passing over to pixel-based methods, a mention should be done on a *human-vision-driven* object-based method: photointerpretation. It was developed together with the use of aerial photography as a means for geological and forestry surveys, and is based in the visual differences that different landcover types yield. The differences are exploited by the photointerpreter in order to divide the image into homogeneous regions that usually are later classified with the aid of a limited field survey. Visual interpretation of EO data is still common practice, since up to now, there is no computer program able to emulate the perceptual and abstracting capabilities of humans. However, humans show some serious limitations in order to fully exploit latent information of data.

First, the number of grey levels distinguishable to the human eye (say some 16) is considerably smaller than the dynamic range of most EO data. Similarly, humans can only compare three bands simultaneously (RGB colour composites). Yet the main drawback is that visual interpretation is based on subjective judgements, as for example where to draw a sharp boundary between two patches that blend gradually into one another. Actually the subjectivity of boundary placement is the major factor contributing to positional error (Green & Hartley 2000c) of the produced thematic map. This inconsistency may cause serious problems in the updating even if it is done by the same person (Ahlcrona 1995), making unreliable any conclusion about changes in landcover drawn with this method.

These shortcomings, together with the slowness, high cost and scarcity of skilled interpreters, have directed the research effort towards automated methods. Given the multidimensional nature of most EO data and their limited spatial resolution, spectral patterns have been favoured against spatial ones, so that pattern recognition in Remote Sensing is seen exclusively as a waveform discrimination problem (see e.g. (Fukunaga 1972)). With this view, the signatures of individual pixels became soon the undisputed basic units of the analysis. The dark side of this approach, so dark that many practitioners are still unaware of it, is that the latent information embedded in the spatial structure of the image is almost completely ignored. A brief description of pixel-based image classification is given in the next section, including a systematic critique.

1.11. A critique to pixel-based image classification

In every scientific community there is a general accepted framework, constituted by formal theories and trusted methods, that shapes customary work (Kuhn 1962). In the case of the community of practitioners of RS digital image analysis, such framework originated from the older tradition of *reflectance and emissive spectrometry*. The latter consists of a series of instrumental techniques developed in chemistry and physics for determining the composition and other properties of materials, from organic compounds to stars. Such framework, that seizes the multiband nature of RS imagery, will be called hereafter the *spectrometric approach* (SMA). SMA is what in RS literature is called *quantitative analysis*. Since most 'quantitative' methods are based in one way or another on waveform discrimination, I prefer the former name. Although waveforms are not necessarily spectral, they are analogous to spectral signatures, hence the name.

The spectrometric approach uses pattern recognition methods to group individual waveforms, or signatures, into classes. A signature consists in an *n*-component vector where each component usually is the value taken by a given individual pixel in each of *n* bands. Hence the basic (areal) unit of the analysis is the individual pixel. Pixel-based classification is so widespread (as matter of fact, it is practically the only classification featured in most commercial RS image analysis products) that it constitutes the prevailing paradigm of this discipline. In this section I discuss the inappropriateness of using individual pixels as the basic units of the analysis, in the context of landcover mapping. In general, the validity of the spectrometric approach is not questioned here, although it will be addressed in the last section of chapter 2.

Image classification, as conceived by the spectrometric approach, is the process of delineating the regions of the multidimensional data space associated with each class of interest w_i (i=1,..., M). The data space is populated by signatures that have to be allocated to some user-defined class. In the pixel-based paradigm, each signature comes from a single pixel. The set of classes must meet simultaneously three conditions (Landgrebe 1999):

- Exhaustiveness: there must be a class to assign to each pixel in the image, i.e. there can be no unclassified pixels.
- Separateness: the classes must be separable to an adequate degree in the data space.
- Usefulness: the classes must be of informational value for users and meet their needs.

The separability constraint requires that, as a general rule, signatures from different classes are relatively distant from each other in the data space. In other words, if a signature is to be correctly classified, then the majority of its nearest neighbours in the data space should belong to the same class, i.e. it should have a *pure neighbourhood*. Conversely, when two or more classes occupy the same tracts of the data space, i.e. when they overlap, the signatures populating these tracts will have a *mixed neighbourhood* and therefore it is likely that they are confused. Therefore the degree of overlap between two given classes will account inversely for their separability (Schowengerdt 1997).

The classification is carried out with the aid of a set of *discriminant functions* g_i (one for each of M classes), such that given a signature X, g_i (X) is greater than the other g_j when X belong to class w_i. In other words, X is classified as a member of class w_i if and only if

g_i (X) $\geq$ g_j (X) for all j=1,2,... M (Landgrebe, 1999).

In order to build the discriminant functions, some amount of ground information (obtained e.g. by field surveys), showing the 'true' class of known locations within the scene, must be available beforehand. An exception to this occurs when the classes are identified according solely to the clustering patterns found in the data space. In this case, known as *unsupervised classification*, the result is a partitioning of the data space into *spectral classes* that are assigned *a posteriori* to *information classes* (those appearing in the map legend) by an analyst based on *ad hoc* collected ground information. The number of *clusters*, or spectral classes, must be limited by the analyst, and the algorithm employed is usually aimed at minimising the variance within each cluster, that is, maximising the variance between clusters[1].

Conversely, common *supervised classification* methods use a set of *training pixels* from *a priori* known locations to estimate the multidimensional probability density function associated with each class, which is used as the discriminant function for that class. Then the most likely class for each pixel is selected, leading to a minimum average classification error given the estimated functions. This process is most simple (and this is why is so popular) for the case in which classes are *normally* distributed (as in fact it is often assumed), only involving the calculation of class mean vectors and covariance matrices (Landgrebe, 1999).

In cases where the gaussian assumption is not suitable because some of the classes show a multimodal distribution (i.e. consists of several distant clusters), a method that has become a popular choice is an *artificial neural network* (ANN), where the training pixels are used

[1] Recall that the total variance of a data set consisting of disjoint groups of data is the sum of the internal variance of the groups plus the inter-group variance.

empirically to construct the discriminant functions. A typical classification conducted with an ANN could be as follows. First, the parametric form (either linear or non-linear) of the discriminant function is chosen, and then the weights of each band are set to an arbitrary initial value. Later these values are adjusted iteratively, with the aid of the training pixels, as to minimize an error function. When the error is adequately small, the training process is stopped and the whole image is classified. By using an ANN one can achieve better accuracies than with statistical methods, at the cost of not knowing what has being done. This black-box problem (Benitez, Castro, & Requena 1997), together with a) the unstable balance between minimising the error function and avoiding excessive training (overfitting); b) the lack of theoretical basis; and c) the lack of definite criteria for choosing the best ANN architecture for a given task; has generated a controversy on the use of ANNs as a means for image classification (Egmont-Petersen, de Ridder, & Handels 2002).

The dependence of the result on training data is not exclusive to ANN implementations. The accuracy of supervised classifications depends heavily on the quality of the training data, even more than on the actual classifier used (Buttner, Hajos, & Korandi 1989). Moreover, the same classifier can produce different results on the same image when trained with a different data set (Smits, Dellepiane, & Schowengerdt 1999). As a consequence, the result is prone to reflect inconsistencies in the selection of training samples. Thus 'good' training data must be fully representative of the respective class, so that they should be well distributed across the scene, and at the same time they should constitute a homogeneous sample of the class[1]. Since both objectives conflict, careful selection of training data is an arduous task. This problem is amplified in the case of hyperspectral data, since the number of training samples required to define the classes quantitatively grows very rapidly with the number of bands (Landgrebe, 1999).

The accuracy of the classification is measured using a contingency table *or confusion matrix* that compares for each landcover class the predicted class with the actual one on the ground, computed from a subset of signatures from known areas that were not used for training. There are a number of methods to measure accuracy from this table (Janssen & van der Wel 1994;Stehman & Czaplewski 1998), the simplest being the percent correctly classified,

[1] The homogeneity constraint is partially solved by picking up blocks rather than single pixels, because near pixels are usually correlated and therefore will show similar values. Collecting training data this way is fast and easy, but it violates the independency assumption of statistical sampling (another thing most practitioners are unaware).

usually called itself '*accuracy*'. In general it ranges from 60 to 85 per cent, which is normally below the user requirements (Lins 1994). Another inconveniences are the often biased spatial distribution of errors, and the significant differences in error rates that are usually found among the classes (Davis & Simonett 1991). The lack of spatial consistency, manifested by the mottled and noisy appearance of classified images, hinders seriously the integration of the result into a GIS when the latter requires the *vectorization* of the classified image. Therefore, in many cases the output image must be post-processed in order to repair such inconsistencies. In doing so, one is implicitly acknowledging a partial failure of the SMA when the classified units consist of individual pixels.

Post-classification techniques take into account *spatial context*, i.e. the relationships between pixels in a neighbourhood, to improve the spatial consistency of per-pixel classifications. Examples are the various morphological filters applied to classified images, or the process of probabilistic label relaxation (Richards 1993). The neighbourhood of each pixel is usually defined by a 3x3 or 5x5 pixel matrix, or *window*, centred at the pixel. However, the assumption that the context of a pixel relies only on its first order neighbours can be increasingly restrictive with image resolution. On the other hand, enlarging the window increases the risk of including pixels belonging to a different patch and decreases the spatial accuracy of the result.

Confusion increases when the pixel size is close to the mean size of the *objects of interest* (landcover patches), leading to a high proportion of *mixed* pixels located between adjacent patches, which are prone to be misclassified. When pixel size is much smaller than the patches, the proportion of mixed pixels will be negligible, but in turn the spectral variability of the patches will increase, the rate depending on landcover type, causing further classification problems (Hsieh & Lee 2000). The failure of the spectrometric approach in very high resolution (< 5m) data is partly due to the non fulfilment of a implicit assumption not acknowledged by many practitioners: pixel size should be large enough as to include a sufficient number of elements producing a typical signature of the class (Woodcock & Strahler 1987).

The former assumption can be restated using Goodchild's (1994) notion of **spatial resolution of a classification**. The latter may be defined as **the minimum size of the circle, expressed by its diameter, over which the surroundings of a geographic point have to be observed**

in order to define the label at that point. This circle will be called hereafter *classification disk*. If the classification is based on the presence and particular arrangement of some individual entities such as trees or buildings, it is obvious that the area of the disk should exceed substantially the extent of those (sub) objects, so that a sufficient number of them is included in the observation. Then the assumption requires that the spatial resolution of the imagery is coarser than the one of the classification scheme. Note that the minimum diameter is class-dependent, increasing with the size and spacing of the subobjects defining each class, e.g. grassland (say 1 meter) – dense forest (5-10 m) – sparse woodland (30-50 m) – urban (100-200 m). Therefore the spatial resolution of the overall classification scheme will be the diameter of the maximum minimal disk (the one corresponding to the urban class in the example).

When spatial resolution of an image is coarser or close to that of the classification, (big) patches of all classes will likely appear as smooth surfaces. As the pixel size decreases, patches of some classes will begin to show an increasingly coarse texture, up to a point where their constituents become resolved, and therefore the spectral response of the sensor is more readily related to the properties of these individual entities than to the ones of the patch. Instead of accepting the fact that the spectrometric approach is not appropriate when applied to individual pixels of images of higher resolution than the classification, the community of practitioners, unconsciously willing to force nature into the box supplied by the paradigm, has either simply ignored the problem or patched it up with some post-classification technique.

In summary, two basic assumptions underlie the spectrometric approach to image classification: **1)** that the piece of terrain from which the measurement is drawn is large enough to include a sufficient number of elements producing a typical response of a landcover class; and **2)** that each landcover class shows a negligible degree of overlap in the data space with the other classes. Now we shall see that both assumptions cannot be fulfilled simultaneously if the spatial distribution of the phenomenon under study is not taken into account, as e.g. it occurs when data units are drawn systematically from a grid.

Suppose we establish a square plot on the ground of the same dimensions than the areal units (pixels) used for classification. Imagine further that we cover its sides with opaque walls, so that an observer standing on it cannot see anything out of the plot. Then, if the first assumption is valid, there should be enough evidence in the plot as to correctly classify it.

Otherwise the hypothetical ground observer would be liable to make the wrong guess, and this would happen more often as the plot size is reduced. Imagine the case of a 1m-side plot (analogous to an Ikonos pixel) that is randomly placed within a sparse mixed woodland. It would be impossible for the observer to identify this piece of terrain as a belonging to a 'sparse mixed woodland', no matter the actual position of the plot. Even if we allow observation in a 3x3 m^2 plot (which would be analogous to contextual –neighbour influenced-classification), the identification would be wrong again.

Now assume that the pixel size is large enough as to fulfil **1)**, and lets explore **2)**. This assumption has two underlying premises: *a*) that there exist natural groups, or clusters, within the data space; *b*) that each cluster has a prevailing landcover class, i.e. that most of the signatures compounding it belong to the same class.

Premise *a* presupposes a non-uniform arrangement of signatures in the data space. This can be taken for granted, given the patterned structure of the imaged territory. A different question is the discovery of natural groups within the data space, since the concept of 'natural' is vague. Let us simply assume that naturalness is gradual and it refers to the existence of discontinuities, or quasi-empty space, separating the clusters. In the absence of such discontinuities, signatures close to the boundaries separating classes would have *mixed neighbourhoods*, and therefore would be liable of misclassification. Premise *b* is at the core of the spectrometric approach, for if there were clusters densely and randomly populated by signatures of different classes, the separability of the classes involved would be seriously compromised.

Now lets study the case of *mixed pixels*, i.e. pixels lying on the boundary between adjacent patches of different class in the image. Their signatures consist of a mixture of typical signatures from two or more classes. Therefore, they are located in the region of the data space separating clusters of these classes, which in turn is supposed to be constituted of quasi-empty space. This circumstance requires the proportion of mixed pixels to be negligible. But if assumption **1)** is fulfilled, the only way to have a negligible proportion of mixed pixels is that the spatial configuration of the territory is a simple one, consisting of big patches (hundreds of times bigger than the pixel) with convex shapes (so that the edge density is low). Such configuration is the exception rather than the norm, since most landscapes can be conceived as mosaics of fragmented, intermingled landcover types. Therefore, for most situations, assumption **2)** cannot hold if the first one is fulfilled.

The only situation in which both assumptions may hold is for a reduced set of broad classes that are spatially segregated over the territory (i.e. there are no holes of a different class within the patches). Since, as explained in 1.4, the more we generalise in geographic space, the larger the objects of interest, it can be expected that patches corresponding to broad categories are big enough as to keep at bay the number of mixed pixels. In any case, an approach that is only suitable for a few restricted situations should not be taken as a general paradigm, and this is the case of pixel-based classification, which is used irrespectively of the pixel size, the number and type of classes, and the nature of the territory.

The problems of pixel-based classification are known since long ago (Markham & Townshend 1981;Woodcock & Strahler 1987). Basically, when the pixel size is below a certain threshold (given by the spatial resolution of the classification), intraclass variability is high, as well as class overlap, hindering separability. As the pixel size is enlarged, class variability is reduced at the expense of increasing the number of mixed pixels, and therefore decreasing the accuracy. The only novelty here is the perspective, consisting of two nested paradigms and their underlying assumptions, under which this problem has been re-examined. The wider paradigm conceives the territory as made of distinct homogeneous materials (landcover classes) with unique spectral properties that are spatially distributed into disjoint pieces larger than a pixel. The narrower paradigm establishes an exhaustive systematic (regularly distributed) sampling scheme as the means to retrieve the spatial distribution of the classes.

The question is not one of determining whether pixel-based analysis is invalid or not, since there is no dichotomy from inadequacy to suitability but gradation. The point is that if we stick to pixel-based classification, we are never going to achieve the level of accuracy required by sound landcover mapping, since the basic assumptions of the spectrometric approach cannot be fulfilled simultaneously if the basic unit of the analysis is the individual pixel. It could be argued that the problem of mixed pixels could be tackled with spectral unmixing techniques, and I would not deny it. The problem is that the trend in RS imagery is towards higher spatial resolution, and then we encounter a hard nut to crack: the spatial resolution of the classification.

In order to make the spectrometric approach temporarily compatible with a soft version of the object-oriented paradigm, let us now imagine that the measurements are not drawn from the cells of a regular grid superimposed on the territory but from a set of jointly exhaustive, mutually disjoint, irregularly shaped cells, or *segments*, forming a partition of the image. Suppose further that i) such partition keeps a good correspondence with the structure of the territory, so that the boundaries separating the segments correspond to discontinuities on the latter; ii) all the segments exceed the minimum size imposed by users as to be representable and potentially meaningful for them; and iii) all the segments are relatively homogeneous, so that their degree of homogeneity is higher than the one that would have the union of any given segment with any of its neighbours.

This situation would be equivalent to an imaginary sensor consisting of detectors that are able to adapt their IFOV to the structure of the imaged scene, so that it deforms and expands the shape of the IFOV until it finds a discontinuity, regularising the response of the surface bounded by such discontinuities. In this hypothetical situation, each resolution cell would correspond to a segment instead of a square pixel. Hence the data volume would be considerably reduced, consisting only in one signature per segment. Since in principle there are no 'mixed segments', and the segments (in virtue of ii) are larger than the classification disk, assumptions 1) and 2) of the spectrometric approach can hold simultaneously under an incomplete version of the *object-oriented* paradigm that uses the same class concepts than the SMA.

The term 'object' is used to refer to a region of the image having a unitary and cohesive identity that is closely related to the one of the geographic object that gave raise to such structure in the image. Once the object is identified and contextualised, the relation between the object and the region of the space occupied by the object is no longer one of identity, since objects may move, grow, shrink or even disappear, but regions by necessity are located where they are and have the extension and shape they have (see 2.2.7). Hence the term *object* is best suited than *segment* or *region* for a diachronic study of the territory. Perhaps 'zone-based' would be the best term, since *zone* (a region distinguished from adjacent parts by a distinctive characteristic) captures better what the partition consists of, but it is important to stress that the thesis is intended to contribute to an ongoing effort to change the way RS images are analyzed for landcover mapping: the shift towards object orientation.

1.12. The quest for object orientation in Remote Sensing

Customary analysis of RS images is based on the utilities built in commercial software like Erdas, Envi, ER-Mapper, and so on. The classification methods they provide, except for a few exceptions, rely on the spectrometric approach developed in the 70s. The common characteristic of all these methods is that they hardly exploit the spatial structure of the images. Their output is commonly considered unsatisfactory from an operational point of view, especially when the task is directed towards the maintenance of geographic databases within a GIS environment. As a result, RS products are not as frequently used for natural resource monitoring and management as envisioned by the engineers who developed them. Consequently the demand for RS images is overwhelmingly lower than the current supply. In other words, the capacity to acquire data exceeds by far the capacity to produce information from the data. This confronts space agencies and vendors with a serious problem. The former find difficult to justify expensive investment in Earth Observation programmes and desperately search for new users and applications. The latter sadly confirm one year after another that the pace of sales do not follow their expectations.

The response to crisis usually takes the form of a paradigm shift that triggers new developments (Khun 1962). In Remote Sensing, it can be viewed as a shift from the spectrometric approach to the object-oriented approach. The latter uses objects in addition to classes in order to model the landscape. Generally speaking, an object is anything to which a concept applies. More specifically, an object represents an individual, unit, or entity, either real or abstract, with a well-defined role in the problem domain (Booch 1991). Vg a landscape object is a patch, defined as a discrete spatial unit having a certain minimum extension and differing from its surroundings in nature or appearance. A quite different concept is a software object, which is a code module that wields data and the code that manipulates the data into a single entity. However both have in common that any single object is an instance of a class. A class is a set of objects that share a common structure and a common behaviour. The class relations among objects are represented in 'kind –of' hierarchies (taxonomies) that provide inheritance, and structural relations among objects are represented in 'part-of' hierarchies (partonomies) that provide encapsulation (information hiding).

The object-oriented approach is especially useful for representing and interpreting the enduring structures of landscape, integrating relevant physical entities (patches) into a coherent relational framework. In order to achieve such framework, two changes are needed:

on the one hand, the shift from individual pixels to image objects (groups of connected pixels that potentially correspond to patches) as the basic units of the analysis, and on the other hand, the shift from classes conceived as types of homogeneous materials to classes referring to types of landscape objects. This section is a brief and incomplete account on the path already followed towards the first shift. The second one, which has hardly been explored up to now, is addressed throughout Chapter 2, and its implications discussed in section 2.7.

The textured appearance of RS images triggered early attempts to incorporate spatial concepts in their analysis, since texture is a disturbing factor for the spectrometric approach. Spectral classifiers were originally developed for Landsat MSS images at 80 m resolution, hence the assumption of low local variance (smooth texture) of landcover classes was more plausible then than now. Each pixel was considered as a sample measurement from a larger element (patch) made of a particular homogeneous material (landcover class), and as such it could be analysed independently of neighbouring measurements. Coarse texture areas were assumed to correspond to classes whose constituents have a size in the same range than pixels and therefore cannot be considered homogeneous at the resolution of the image. Note that the alternative interpretation that coarse texture is due to different classes interspersed at intervals close to the pixel size was discarded in terms of explanation but it was implicitly accepted as an anomaly referred to as 'scene noise'. One possible solution to tackle coarse texture is to smooth the image with a filter at the expense of decreasing spatial resolution. If by any reason the original resolution is to be retained, then the only alternative is to divide the scene into regions defined as groups of contiguous pixels which may be presumed to belong to a common class and to extract average signatures from these regions.

The easiest way of doing this is when there is a previously existing partition that can be assumed to reflect adequately the spatial distribution of classes within the scene, as e.g. a cadastral vector layer of agricultural fields where there is only one type of crop per field. The mean (and possibly the variance) of pixels within each field is used as the input for classification, and the accuracy of the result is considerably improved. This method, which was proposed as early as in 1969 (Huang 1969), is usually referred to 'per parcel' or 'per field' classification to distinguish it from other methods where the partition has to be defined from the image itself. The latter are based upon image segmentation.

The goal of segmentation is to partition the image into a set of jointly exhaustive, mutually disjoint regions that are more uniform within themselves than when compared to adjacent regions. Segmentation techniques provide a primary model of the spatial structure of the image that is used subsequently to form classified objects. The ECHO classifier developed by David Langrebe's team at Purdue University in the 70s included a segmentation algorithm that can be viewed as the first milestone in the quest towards object orientation in Remote Sensing. The image was divided into 'cells' of 2x2 pixels that were subject to a simple test of statistical homogeneity. Cells failing the test were assumed to overlap a boundary and were later classified on a per-pixel basis. Adjacent cells passing the test were selected and subsequently subject to hypothesis testing for statistical similarity. Cells found similar were merged or annexed into regions (note that although the image is processed in a single pass, similarity is considered transitive, and hence regions can be formed by strings of merged cells). 'In this way an object can grow to its natural boundaries, whereupon either the cell selection or annexation test will fail' (Landgrebe 1980).

Although Landgrebe made available a Fortran implementation of ECHO (user's guide included) to the community of practitioners, it seems it raised little interest. On the one hand, the homogeneity and similarity thresholds had to be set on a trial and error basis. On the other, the merging sequence could lead to unwanted results, since it allows the creation of regions with large differences between pixels at opposite extremes. In the end, I guess that most people thought that similar results could be achieved by post-processing the pixel-wise classified image, therefore they did not bother checking out the new method or developing a similar approach. The lack of interest in image segmentation is confirmed by the fact that only a handful of papers related to segmentation of RS images were published in the 80s (e.g. (Nazif & Levine 1984;Derin & Cole 1986;Cross, Mason, & Dury 1988)). As a point of fact, classical RS textbooks (e.g. (Richards 1993)) included no section on the subject.

A contemporary survey by Haralick (Haralick & Saphiro 1985) classified existing segmentation algorithms in three classes: clustering, region growing and split-and-merge procedures. The second category was further subdivided into single linkage (proper region growing), hybrid linkage (edge detection) and centroid linkage (region merging). This classification is somehow confusing, since e.g. many region-merging methods can also be viewed as spatially constrained clustering methods (see below). However it is important to note that region merging (aggregation of adjacent regions) and region growing (annexation of individual pixels to neighbouring regions previously formed from seeds) are frequently used as synonyms. Notwithstanding it, the term 'region merging' reflects better what the methods studied here do, and hence will be preferred.

Another major contribution to the field is the stepwise optimisation algorithm of Beaulieu and Goldberg (Beaulieu & Goldberg 1989). It begins by considering single pixels as the initial regions. At each iteration, two adjacent regions are merged provided they minimise a heterogeneity criterion. The candidate pair that produces the least increment in heterogeneity (i.e. that shows the highest degree of fitting) is merged first. The initial regions are merged gradually in this way. Hence the process can be seen as a hierarchical agglomerative clustering constrained to adjacent regions. The height in the dendrogram of each merger is the value of the heterogeneity criterion for that pair. Partitions at higher levels of the dendrogram consist of fewer regions, so that this sequence of partitions may reflect the hierarchical structure of the image. Each partition is optimal regarding the minimization of the heterogeneity criterion, but the procedure is too slow, since it allows only one merge per pass (a strategy known as *global mutual best fitting*). Besides, it leads to an uneven growth of regions between areas of smooth and coarse texture.

Woodcock and Harward (Woodcock & Harvard V.J 1992) introduced a faster algorithm that allows multiple mergences per pass. Given a target global threshold T_{glob} of a dissimilarity measure (analogous to the foregoing heterogeneity criterion), they set a series of intermediate pass thresholds T_{pass} ($< T_{glob}$) of increasing value. Then at each pass, two adjacent regions are allowed to merge if and only if: 1) neither regions has previously merged on this pass; 2) the distance between the regions is less than T_{pass}; and 3) each region is the most similar neighbour of the other (*local mutual best fitting* criterion). Note that condition 1) is only necessary for cases in which there are ties in the dissimilarity distance and thus there are more than one nearest neighbour. As these authors noted, the global threshold alone leads to inadequate results, since usually there is a great disparity in size of the output regions. Areas marked by coarse texture will consist of many small regions (often individual pixels), whereas smooth uniform areas will be segmented into a single large region. Therefore they supplemented their algorithm with some size constraints that prevented excessive growing in smooth areas and forced the development of regions exceeding the minimum size of the mapping unit in areas with high local variance.

The most important contribution of Woodcock and Harward's (1992) paper to object orientation in Remote Sensing, rather than this algorithm[1], is their nested-hierarchical scene model. This model assumes that the spatial structure of RS images reflects a hierarchy of nested levels, in which each object (e.g. a stand) can be conceived simultaneously as a whole made of smaller wholes (e.g. trees) and as a part of a larger whole (e.g. a forest). They pointed out that the piecewise homogeneous model underlying traditional segmentation methods is unrealistically simple, since it assumes that the objects of interest (landcover patches), as manifest in images, have internal variances that are both low and equal. Such assumption is inadequate for RS imagery, since different landcover types exhibit differing levels of internal variance given a fixed pixel size. As a result it is very unlikely that all the regions defined by a conventional segmentation method correspond to patches of the same level of the landscape hierarchy. Hence the need for size constraints focusing on a certain hierarchic level. This model is the counterpart in Remote Sensing of the hierarchical patch dynamics paradigm (Wu & Loucks 1995) in Landscape Ecology: "... The attempt here is to tie the hierarchical structure of images to the hierarchical nature of landscapes/classification schemes, and to note when this relation breaks down and why" (Woodcock & Harward 1992).

Turning back to segmentation algorithms, the problem of beginning the region merging process by individual pixels was twofold. On the one hand, it was computationally expensive, and on the other, it precluded the usage of statistical dissimilarity measures as long as there were 1-pixel regions left. One of the first attempts to tackle this problem is the work of Lobo (Lobo 1997). Prior to the region merging stage, he applied an iterative edge preserving smoothing (EPS) due to Nagao and Matsuyama (Nagao & Matsuyama 1979). The output of the filter is a piecewise constant image made of tiny segments, where each segment represents a homogeneous region in the original image, darker or brighter than its surroundings (i.e. a blob). After labelling each segment, the resulting image can be taken as the baseline partition for the region merging stage. This partition can also be conceived as a primal sketch of the image, where the blobs conforming its spatial structure are formalized into segments. Although the filter he used was rather rudimentary (it was based on directional masks around each pixel that produced segments of biased round shape), the procedure contains the germ of object orientation, since these primal segments can be viewed as the basic objects that

[1] Actually this algorithm was not the first of its kind to appear. Vg the Swedes (Hagner 1990) were already using a similar method for automated forest stand delineation, which used another dissimilarity measure and did not apply the mutual optimality criterion. Unfortunately Hagner did not publish his work in an international journal, and hence it can be assumed that it was unknown to most foreign researchers.

compound the image. Unfortunately, Lobo (1997) deviated from the object-oriented path in the next stage of his method, by considering these segments not as objects themselves but as collections of measurements drawn from an larger object (patch) made of a particular homogeneous (with gaussian distributed properties) material (landcover class). Hence he chose a statistical approach to carry out the subsequent region merging process, which he named Iterative Mutually Optimum Region Merging (IMORM).

IMORM, likewise Woodcock and Harward (1992) segmentation, followed the local mutual best fitting criterion and the stepwise increase of the dissimilarity threshold, but imposed no size constraint, since it assumed a piecewise homogeneous model of the scene. That is to say that IMORM considers the set of pixels within a segment as a sample from a bigger stochastic population characterized by a gaussian distribution. Hence the goal of the region merging stage was to retrieve maximal sets of spatially connected samples belonging to the same population (class). The way to achieve such goal was to test the null hypothesis that the samples extracted from two adjacent segments are in fact observations of the same population. Note that the overall population is the set of pixels that belong to that class. In this sense, statistical segmentation relies both on the spectrometric and pixel-based paradigms.

The dissimilarity measure used by IMORM was the normalized difference between two means of samples of normal distributions, or *t-ratio*[1]. If for a given iteration and candidate pair (adjacent segments which are the most similar neighbour of each other), the null hypothesis could not be rejected at a confidence level given by the current t-ratio threshold, both segments were merged. Although apparently sound, this approach reveals an inherent contradiction. The t-ratio measures the statistical significance of the difference between both samples rather than the magnitude of that difference. Therefore when the temporary t-ratio threshold is increased after an iteration, candidate pairs that formerly were considered significantly different may be merged. Such incongruity can be mitigated if the final threshold is low enough as to avoid undersegmentation, i.e. the merging of segments of different class. In any case, the conclusion is that the selection of a dissimilarity measure for early segmentation is not a trivial question.

The classification stage of Lobo's (1997) method returned to object orientation. The final segments were considered as patches (or part of patches) that were to be classified through a

[1] t-ratio is the absolute value of Student t test. Again, Olle Hagner (1990) was the first in using this approach, but his work remained unknown for most non Scandinavian researchers.

linear canonical discriminant analysis. Each segment was treated as a basic unit defined by a set of attributes in a similar fashion as individual pixels are treated in pixel-based spectrometric methods, differing in that textural attributes are more efficiently extracted from segments than from kernels. However he did not use any spatial (size, shape) or contextual (relations between neighbouring segments) attributes. The problem is that contextual information tends to be declarative in nature (represented symbolically and independently of the methods to perform inferences on it), so that it is not trivial how to incorporate it into an algorithm.

Early attempts (Goldberg, Goodenough, & Plunkett 1988;Moller-Jensen 1990;Ton, Sticklen, & Jain 1991) to analyse RS images using a knowledge-based framework (i.e. interpretation-guided segmentation) did not result in any operational method, with perhaps a couple of exceptions. One of these was the System of Experts for Intelligent Data Management (SEIDAM, http://www.aft.pfc.forestry.ca/seidam_e.html), a Canadian prototype system of multiple expert systems (developed in the 90s and apparently no longer used now) that could update existing forest resource inventories and respond to queries by dynamically selecting RS data in a distributed GIS environment. The other is AIDA (Tönjes et al. 1999), a knowledge-based system for the interpretation of RS data, which was the first system using a semantic net to formulate knowledge about scene objects. Semantic nets (Brachman 1977) are directed acyclic graphs where the nodes represent the objects of interest (including subobjects and superobjects) and the links form the relations between objects. The objects properties are described by attributes attached to the nodes. They have a value measured from the data and a range describing the expected attribute value for each type of object. Some of the attributes may be relational, taking e.g. into account topological relations that may affect the attributes of neighbouring nodes. The instantiation (allocation of an image object to a predefined type of object) of objects is conducted by a judgement function that computes the compatibility of the measured value with each hypothesis regarding the type of object or subobject each image object is. Finally, an inference engine determines the sequence of rule execution, which is based on a model-driven interpretation with a data-driven verification of hypotheses.

However, the real break-through in object-oriented analysis of RS images, that for the first time provided users with an operational tool, was the introduction of the eCognition software at the ISPRS Conference in Amsterdam in summer 2000. eCognition (http://www.definiens-imaging.com) is an analysis-specific (no image preprocessing built in) program for multiresolution segmentation and object-oriented fuzzy-rule classification, specially suited for very high resolution imagery. It is based upon the Fractal Net Evolution Approach (FNEA) derived from the ideas of the Nobel laureate Gerd Binnig (Klenk, Binnig, & Schmidt 2000). FNEA describes complex image semantics within self-organizing hierarchical networks, in which the structure of each level is similar to the one of the others (hence the 'fractal' adjective). Objects derived from the input image change their states (structure and meaning) stepwise according to contextual influences and converge to a semantically coherent hierarchical arrangement through alternating procedures of segmentation and classification (hence the name 'evolution').

The construction of the hierarchical semantic network of image objects is based initially upon a multiresolution segmentation (Baatz & Schape 2000). A typical analysis may consist in producing two (or more) coordinated partitions, one fine grained (with many small regions of relatively uniform size) and another coarse grained (with larger regions whose mean size exceed the one of the former partition by orders of magnitude), where the boundaries of the coarser partition are formed by boundaries existing in the finer one. Alternatively, the coarser partition can be obtained from a pre-existing GIS vector layer (if e.g. the aim of the analysis is map updating). Each region is considered an image object whose projection on the ground may have a definite meaning for users, either as an entity of its own or as a part of larger structure. Membership of an image object to a class is evaluated by fuzzy membership functions (one for each attribute) that maps class descriptions into the [0,1] real interval according to the typical values in each attribute showed by sample objects of that class.

The set of classes and attributes may differ between partitions, with classes corresponding to a higher level of abstraction for the coarser partition. The overall classification scheme is structured in three kinds of hierarchies. In the *inheritance hierarchy*, class descriptions defined in parent classes are selectively passed down to their child classes, reducing redundancy and complexity in the class descriptions. In the *group hierarchy*, classes are combined into classes of superior semantic meaning. A class may be part of more than one group, and the grouping of classes may not coincide with the structure of the inheritance hierarchy. Finally, the *structure hierarchy* put together different classes that may compound complex heterogeneous geographic objects like a city.

All these semantic relations can be put to work in either a bottom-up or top-down fashion. In the first case, connected image objects of the finer partition that represent identical structures or that are parts of identical structures are merged into a new image object. In the second, the congruency of subobjects compounding a superobject is evaluated and eventually the hierarchical connection between them may be deallocated and assigned to a neighbouring superobject. Since these changes may affect the semantic value of neighbouring objects at the same or higher level of the object hierarchy, the classification is done iteratively in cycles in which each object is classified over and over taking into account the changes in the classification of networked objects. The result of this process is a network of classified image objects with concrete attributes, concrete relations to each other and concrete relations to the defined classes.

The set of initial partitions upon which the classification starts is given by the multiresolution segmentation algorithm due to Baatz and Schape (Baatz & Schape 2000). Likewise Woodcock and Harward's (1992) method, it starts with individual pixels as the initial segments. It also follows the local mutual best fitting criterion, but instead of using an stepwise increase of the dissimilarity threshold, at each iteration it distributes the candidate pairs to be merged (those having a dissimilarity value smaller than the global threshold) as far as possible from each other over the image, and these locations cannot be close to segments that were merged in the previous iteration. In this way, it achieves a uniform growth of segments throughout the image, so that the final segments have all a similar size. Since a conservative (small) threshold permits fewer merges than a greater one, the mean size of segments will grow with the value of the threshold. For this reason it is called the *scale parameter*. A hierarchy of increasingly coarser partitions can be obtained by simply raising the scale parameter. Such hierarchy is better suited than the one of Beaulieu and Goldberg (1989) for the study of the landscape at different levels of generalization, since objects at a given level have all roughly the same scale (size).

Another particularity of this algorithm is the optional inclusion of a form heterogeneity factor in the overall dissimilarity between two adjacent segments of size n_1 and n_2. The latter is measured as the change in heterogeneity produced by their eventual merging, i.e. the difference h_{diff} (weighted by size) between the heterogeneity h_m of the potential merger and the ones h_1 and h_2 of the segments: $h_{diff} = (n_1+n_2)h_m – (n_1h_1+ n_2h_2)$. The overall heterogeneity h is a linear combination of radiometric heterogeneity (expressed by e.g. the mean of the variance in each band of pixels within the segment) and form heterogeneity (expressed e.g. by the ratio between factual edge length and the edge of a square with the same number of pixels than the segment). In this way the segmentation favours the construction of regions with smooth edges and a more or less compact form. Although such approach to tackling the fractal nature of the landscape is conceptually weak, the results are visually appealing. Another inconsistency is the consideration of individual pixels as the initial image objects, since they are artificial units whose shape is in no way related with the spatial distribution of landscape objects. An already proposed alternative (e.g. (Blaschke & Hay 2001)), which will be followed in this thesis, is to detect blobs with some morphological method and use the resulting fine partition as the input for region merging. Morphological segmentation, in contrast to the statistical one, is based on the spatial structure of the image, and has the watershed transform (see 3.7) as its cornerstone. Meyer (2001) gives an unified overview of morphological segmentation, a set of powerful techniques customarily used in computer vision and medical radiology that have been hardly applied to Remote Sensing but that will shape in all likelihood future work in this discipline.

To summarize, efforts directed towards object orientation in Remote Sensing have followed a windy road with many comings and goings. The need for using distinct uniform regions instead of individual pixels was acknowledged since long ago as the only way to tackle the unevenly distributed heterogeneity of landscapes. However, many initiatives got stuck on the piecewise homogeneous model, which neglects the fact that landscape heterogeneity is hierarchically structured. The late acknowledgment of the scale dependency of most landscape attributes and concepts led to the emergence of the hierarchical patch paradigm both in remote sensing (Woodcock & Harvard V.J 1992) and landscape ecology (Wu & Loucks 1995). The natural way of making operational this patch concept is through object-oriented modelling, where each object is a structural-functional unit (a patch) at a given scale that is loosely coupled with both objects of the same level and objects forming part of it or encompassing it. So far, image segmentation seems to be the only operational solution providing a starting point to object-oriented analysis. But most methods have been developed heuristically without a deeper examination of the semantic implications of the segmentation process. As Bittner and Winter (Bittner & Winter 1999) point out, 'for a better understanding of the relationship between objects of the real world and their representations, a better understanding of the underlying ontological and epistemological foundations is necessary'. Chapter 2 aims to contribute to this understanding.

1.13. Motivation, objectives and main contributions

A historical view would help to stress the main points. The first spaceborne sensors were multispectral to compensate for the reduced spatial resolution of the data, with the hope that different landcover types would behave like distinct materials susceptible of being analysed with a spectrometric approach. Hence it was natural to consider each pixel as a sample introduced in a desktop spectrometer, and therefore the individual pixel has been considered the basic unit of the analysis since the beginning of EO. The fact that these samples do not come separately (rather they are knitted into an image full of spatial patterns) was neglected and even considered of little use, except for photointerpretation. Several classification methods were developed based on this approach and soon (even before the launch of Landsat-1) they were established as the trusted common practice, becoming the accepted paradigm in the analysis of EO data.

The initial resolution of these data (80 m) was compatible with the spatial resolution of most classification schemes (see 1.11), but as the technical developments enabled smaller pixel sizes, the radiometric variability of surface features increased. Therefore there was a need to incorporate simple spatial characteristics as adjuncts to the spectral ones (Landgrebe 1980), and textural features were the most successful candidates, enhancing somehow the accuracy of classification (Haralick 1979). This relative success, together with the use of smoothing filters that got rid of the (wrongly called) 'scene noise', allowed the paradigm to survive by tacitly forgetting the basic premise about the spatial resolution. Other approaches pointing towards object orientation (image segmentation followed by segment classification) were suggested as early as 1976 (Kettig & Landgrebe 1976), but not surprisingly they remained ignored until recently. On the one hand, they were outside the paradigm frame, and on the other, they were rightly criticised for several reasons, among others, the dependence on seed pixels and/or merging sequence, the need for used-defined parameters, and the lack of theoretical basis.

The advent, in the beginning of this century, of civilian very high resolution multispectral satellites of the like of Ikonos and Quickbird, has brought into a sharp relief the inadequacy of pixel based analysis when the pixel size is smaller than the classification disk. Neither does texture play the same role than in Landsat-like imagery, for the unresolved elements are in this case leaves or tiles rather than tree crowns or roofs. Simultaneously, segmentation techniques for grey-level images improved significantly as a result of research efforts in the fields of computer vision and medical radiology.

However, these results have hardly been transferred to remote sensing, mainly due to the lack of definite shape (and even crisp boundaries) of the objects of interest, and to the multiband and multiscale nature of the images (Schiewe, Tufte, & Ehlers 2001). As a point of fact, up to now there is only a commercial software (eCognition) devoted to object-oriented image analysis. Although most users have been impressed by its results, the multiscale segmentation algorithm embedded in this software lacks an explicit theoretical framework, and the users have to find useful segmentation levels in a trial and error basis (Blaschke & Hay 2001). On the other hand, there has been little progress in segmentation of colour or multiband imagery (Kartikeyan, Sarkar, & Majumder 1998), due the relative lack of interest of the remote sensing community and to the monochrome nature of most imagery from the other disciplines.

The more practitioners are aware of the inconsistencies of spectrometric methods, the more urgent the need for a new paradigm. The emerging object-oriented paradigm has being around for a good while, but in order to achieve a full conversion towards it, the concept of class should also be altered in order to make it compatible with the hierarchical patch model. To the best of my knowledge, no one has already put forward a solid conceptual basis for the new paradigm, neither given an explicit account on the implications of class-concepts in the analysis of RS images for landcover mapping. By shifting the concept of classes towards types of geographic objects, a new key concept appears as the basic unit of the analysis: the granule, or basic mappable zone. Finally, a general method of segmentation should be developed in order to change once and for all the way users analyse RS images for landcover mapping. The ultimate goal of this thesis is to make a significant contribution to this shift.

Keeping in with the last statement, the purpose of this thesis is threefold:

i) To expose the inadequacy of the spectrometric approach to image classification as a means for landcover mapping (sections 1.11 and 2.7).

ii) To construct a conceptual framework for an alternative approach that seizes the spatial structure of the image and that is based on basic mappable zones, or *granules* (chapter 2).

iii) To develop an implement a tentative version of a general automated method to derive from a multiband image a partition consisting of granules (Chapter 3).

The main contributions of this thesis, listed by objective, can be summarized as follows:

1) Related to the first objective:

- A re-examination of the main critiques to pixel-based image classification under a kuhnian (Kuhn 1962) perspective, in which the former is viewed as nested on a wider paradigm, the spectrometric approach, which is based on waveform discrimination.
- The identification of Goodchild's (1994) notion of spatial resolution of a classification as the key concept to understand of the failure of pixel-based methods in high resolution imagery.
- The exposure of the conceptual incompatibility of spectrometric methods with the object-oriented paradigm when the latter is nested within the hierarchical patch model.

2) Related to the second objective:

- The identification of the need to shift class concepts from types of materials to types of geographic objects in order to make compatible the classification process with the hierarchical patch model.
- The application Thom's (1975) theory of attractors to the morphology of geographic fields. In particular, the identification of this theory as a suitable conceptual basis to define a primal partition of an image.
- The introduction of the *granule* (a region different from its surroundings and larger than the MMU size) as the basic unit for Object-Oriented Classification of RS Images for landcover Mapping (OOCIM).
- The integration of several ontological and epistemological tools into a 3-tiered (commonsensical reality, geographic fields and classified objects) model of landscape in which the last tier is the landcover map. This model is the conceptual basis proposed for OOCIM.

- The identification and exploration of the three basic premises underlying the former model, namely the *coincidence*, *size* and *correspondence* hypotheses. They sustain the plausibility of using RS images for landcover mapping under the object-oriented approach.

3) Related to the third objective:

- The introduction of a new version of non-linear (iterative) diffusion filter, the gradient inverse weighted edge-preserving smoothing (GIWEPS).
- The use of Baraldi & Parmiggiani's (Baraldi & Parmiggiani 1996) normalised vector distance (NVD) to compute a surrogate 'gradient magnitude' of a multiband image.
- The identification of gradient watersheds as the tool with which to apply Thom's (Thom 1975) theory of attractors to image segmentation.
- A novel region merging method: the size constrained region merging (SCRM) algorithm.
- The development and implementation of a general automated method, based on the foregoing achievements, to derive a baseline partition for OOCIM.

CHAPTER 2

Conceptual Foundations

Without generalization, there can be no theory. Without theory, explanation is highly limited if not impossible. If each place is entirely unique then there can be no generalizations, no geographic theory, no anticipation... But if we believe there are commonalities... the door is open for geographic theory.
Grant Ian Thrall, *The stages of GIS reasoning (1995a)*

2.1. Introduction and overview

Maps are models that represent a particular view of geographic reality. Landcover maps partition geographic space into a set of disjoint regions that constitute instances of some landcover types. Each of them is a thematic unit that is portrayed in the map as a *polygon*. The latter represents a piece of land that is supposed to constitute a unitary coherent conceptual entity distinct from its surroundings. But, how are such entities individuated from each other? Drawing boundaries between polygons is at the very heart of thematic mapping, however, little attention is paid by practitioners on the issues that this question arises: Are landcover polygons representations of real geographic objects? And if so, can such objects stand with indeterminate boundaries? Can they survive delineation errors or abrupt changes? How can we delineate polygons based on data (RS images) that are only contingently related to the biophysical properties of landcover? What is the basis to believe in what the image-derived map tells us about the territory? These questions should be answered before any attempt to deal with this kind of maps, since they deeply affect the way they are made, used and updated.

This chapter deals with the conceptual principles underlying landcover mapping. In particular, it aims to explain why we can derive landcover maps from RS images. In doing so, it sets forth the foundations of a particular view of object-oriented analysis of RS images for landcover mapping. Such basis is formalized in a multi-tiered model of geographic reality that ends up in the geographic objects of interest. In the section 2.2, several important concepts are introduced in order to make understandable and contextualize the proposed geographic model. The section elaborates a synthesis of the responses already given by philosophers, psychologists, mathematicians, geographers, and ecologists to issues related to landscape modelling, landcover classification and mapping, and image analysis.

Sections 2.3 and 2.4 explain the model background and motivation. Sections 2.5 and 2.6 deal with the model itself, the first one with an idealistic version and the second with a more realistic version. Finally, section 2.7 discusses the inadequacy of the traditional spectrometric approach in the light of the object-oriented one used here.

2.2. Assumptions, concepts and definitions

Let us depart embracing realism from an ontological point of view (there exist a single physical reality which can be truthfully observed via our *sensorium* or other artificial apparatus, and whose existence is independent of human cognition), and constructivism from an epistemological one (the internal representations of reality are manifold, and those are constructed through concepts reliant on language and individual's experience).

2.2.1. Perceptual constructivism

Constructivism, although a product of contemporary research by Jean Piaget and the Gestalt psychologists *inter alia*, can be dated back to the Kantian notion of *schema* (a mental representation) or even to Plato's doctrine of ideal forms (one recognises real instances of objects by reference to the ideal form). From the constructivist viewpoint, perception is a matter of picking out from the perceived scene, chunks that match some of the pre-defined concepts stored in the observer's knowledge database –her brain, and putting them in the foreground, in an effortless process coined by linguist Leonard Talmy (1996a) *the windowing of attention*. The fact that there is no perception without immediate categorization is also implicit in the work of Piaget (Piaget 1969). An impressive evidence of this need is the case of a man who, recovering his sight after 30 years of blindness, reported:

> "When I could see again, objects literally hurled themselves at me. One of the things a normal person knows from long habit is what not to look at. Things that don't matter, or that confuse, are simply shut out of their seeing minds. I had forgotten this, and tried to see everything at once; consequently I saw almost nothing." (Muenzinger 1942)

Hence, in order to make sense of a scene, the observer has to recognise certain patterns as forming up instances of objects already known and trace over the irrelevant. The easiness with which the perceiver makes such distinctions has evolutionary roots (see (Gibson 1979)), and this evolution has possibly been guided by the MEP principle (see appendix 1): 'perceptual guidance of movements and the movement enhancement of opportunities to perceive better, extend the affordability for animals of energy resources discontinuously located in space and time, therefore expanding the patches of the planet in which energy degradation can take place' (Swenson & Turvey 1991).

2.2.2. Common sense realism

Let us now further assume that all the concepts referring to objects which are susceptible of direct perception and interaction are transparent to the reality beyond, that is to say, the objects to which they refer do exist and have the properties suggested by those concepts. This set of concepts and their relations conform what anthropologist Robin Horton (1982a) calls *primary theory*[1], or what philosopher Barry Smith (1995b) takes as *common sense*, since it is common to all cultures, it is marked by a widespread unforced agreement and it is readily translatable from language to language. Primary theory gives the world a foreground filled with mesoscopic (say between a hundred times as large –e.g. a forest, and a hundred times as small –e.g. a stamen- as human beings), enduring, material objects.

Apart from the thermodynamic reasons mentioned above, there is a good point to believe that our perceived world must be systematically related to the real characteristics of the real world: otherwise, we would not be here. The survival value of perceptual reliability is so overwhelming that if we had not attained it, another creatures having it would have occupied our niche (Campbell 1988). As Barry Smith (2000a) puts it: 'an act of visual perception stands to a visual field as an act of (true) judgment stands to a fact or state of affairs'. In other words, what we have taken for granted here is that the common-sense view of the world is, in certain fundamental features, completely real (Moore 1959).

2.2.3. Prototipicality of categories

Primary theory is structured qualitatively in terms of concepts falling under categories. Each domain (family of categories) is organised hierarchically in the form of a tree, with more general categories at the top and successively more specific categories appearing as we move down each of the various branches (Smith & Mark 2002). The assignment is made according

[1] The term *theory* here refers to a set of definitions which specify the properties and relations of a collection of entities. The adjective *primary* is in contrast to the *secondary theory*, which consists of folk concepts that pertain to what lies beyond the things that are immediately given in perception and action, like religious, magic, pseudo-scientific or psychological beliefs.

to some properties that *typical* members of each category usually have. This means that the internal structure of categories seems to be better characterized by *prototipicality* rather than by an analytic definition specifying necessary and sufficient conditions that all members must fulfil[1]. Thus, a Norway spruce (*Picea abies*), typical of the *tree* category, have more of the properties listed for most trees than a coconut palm (*Cocos nucifera*), and the latter more than a giant cactus (*Carnegiea gigantea*). Hence categories may be conceived as having a radial structure, where a core of prototypes or typical members is surrounded by a penumbra of less typical instances (Rosch 1978). Concepts out of the core may have been added to the category by a *family resemblance*[2] that may have little to do with any objective property, so that the categories cannot be properly modelled by mathematical sets (Lakoff 1987). The empirical discovery of common characteristics would not show that the category in question is not formed up by a family resemblance; what is decisive is the existing practice of explaining the category, which usually consists in a series of typical examples with the rider: 'and other similar things' (Wittgenstein 1999).

2.2.4. The basic level of cognitive categorisation

Another interesting finding of Eleanor Rosch (1978) is that there is a privileged level in the hierarchical tree of categories: the most frequently used or *basic level*, which corresponds to a compromise solution between maximizing informativeness and keeping the number of categories of the level relatively small. For western urban people, this is the level of *fish, tree, table* and *shirt* rather than the more general level of *animal, plant, furniture* or *clothing* and rather than the more specific level of *trout, oak, coffee table* or *dress shirt* (Mark, Smith, & Tversky 1999a). The *height* of the basic level in the categorical *tree* of a given domain will depend on the frequency of interaction with the objects of that domain and the utility of that interaction for the group of people involved. Thus the basic level of the vegetal domain will be further down for foresters than for brokers, and the same will happen to the 'solid water' domain between say Inuits and Massais.

[1] A funny anecdote related to this is one of the many from Diogenes the cynic (s. IV B.C.). Plato had defined Man as a "featherless biped". Diogenes plucked a chiken and brought it into the class with the words "Here is Plato's man." In consequence of which there was added to the definition, "having broad nails".

[2] The *family resemblance* is the cognitive glue that sticks together the members of a category. The term was coined by Wittgenstein in his later work retracting his previous theories, which was published after his death (1951) in his *Philosophical Investigations*.

2.2.5. Taxonomies and Partonomies

Not only knowledge about the world is organised hierarchically in the *taxonomies* (*kind-of* hierarchies) described above, but as explained with some detail in appendix 1, the world itself can be viewed as made up of hierarchies of *part-of* relations or *partonomies* (Tversky & Hemenway 1984), giving way to a granular structure. Every real entity (an object) is made up of parts (another objects) that are obviously smaller than the whole. What is a part of an object at some level of the hierarchy becomes an object itself at a lower level and vice versa, and this relation is not only structural but functional, that is, the existence or activity of the part is reflected in the properties or behaviour of the whole.

A forest may be formed by several stands that contain a number of trees, each of them having many branches within which leaves are nested, and each leaf is made up of cells arranged in tissues, the cells being made of proteins and other molecules, and so on. Some of these molecules may absorb blue or red light and this in turn will yield the green appearance of the forest. At the same time, we do not need to know the emission-absorption spectra of those molecules to acknowledge *colour* as a macroscopic property of the forest. Similarly, we can rely on the colour, tone and texture of the forest canopy to realize from a distance that it is e.g. a pine forest, and if we get a little closer it may suffice for us to distinguish the salmon-red bark of the upper trunk in order to learn they are Scots pines (*Pinus sylvestris*), without having to inspect other parts as needles or cones. In short, the real world shows a hierarchical structure that enables analysis at many levels (what ecologists call *grain*). Through that limited analysis we can classify objects without having to recognize *all* their constituent parts neither know *all* the interactions involved. Then the minimal parts of our analysis simply indicate the level of granularity beneath which we do not care about *what* or *where* the sub-parts are.

Taxonomies and partonomies show two main differences. First, while the former focus on prototypes, the latter are usually employed to separate individuals an their components rather than classes. Second, categories within taxonomies inherit the properties of the superclasses within which they are nested, allowing useful inferences for cognitive economy (if we know a Scots pine is a conifer we can take for granted that it has cones even if we are too distant to see them); partonomies on the other hand do not permit property inferences (Mark, Smith, & Tversky 1999a) (a gap in a forest lacks many of the qualities of the forest; a single tree is not a forest). However, they both have in common the determination of boundaries: between classes in the former case and between individuals (and their parts) in the latter. In both cases, the boundaries are drawn in such a way that:

> **First … the material enclosed within the boundary is felt to constitute a unitary coherent conceptual entity distinct from the material outside the boundary. Second, there seems to be some sense of connectivity throughout the material enclosed within the boundary and, contrariwise, some sense of discontinuity or disjuncture across the boundary between the enclosed and external material. … Third, the various portions of the material within the boundary are felt to be corelevant to each other, whereas the material outside the boundary is not relevant to that within.** [Talmy 1996, p. 240]

2.2.6. Granular partitions

Both taxonomies and partonomies are cognitive devices that articulate reality by systematically imposing boundaries that foreground the objects of interest. Each one can be seen as a system of cells and subcells assembled in a logical tree. The root or maximal cell defines the domain of the partition (e.g. the vegetal kingdom, Europe), whereas the leaves or minimal cells are the 'atoms' of this domain relative to the partition, so that the objects that fit into the minimal cells are treated in the partition as if they had no further parts (e.g. in the Linnean taxonomy, each species would be a minimal cell; in a not-too-detailed political map of Europe, the *lander*, shires, or regions like Bavaria, Yorkshire, Tuscany or Catalonia could be the minimal cells. In between the root and the leaves, there are normally several levels (although there may be none, when the partition is a mere listing), each one with more cells than the former. The tree is structured so that i) each intermediate cell, or *parent cell*, has at least two descendants, or *daughter cells* (so that there are no redundant levels); and ii) there is a unique path from each minimal cell to the root (i.e., Pinus sylvestris cannot be simultaneously Pinaceae and Fagaceae, or Catalonia cannot be part of Spain and France at the same time). Following Bittner and Smith (2001a), such kind of nested cells systems will be called hereafter *granular partitions*.

Smith and Brogaard (2000b) introduced the notion of *granular partition* to tackle the deficiencies of both set theory and mereology (the algebra defined by the *is-part-of* primitive) when used for cognition. The problem with set theory is that sets are identical if and only if they have the same members. If we model *Homo sapiens* as the set of its instances, then this means that humankind becomes a different species every time a person is born or dies. Also set theory allows double counting (overlapping between sets), whereas a sound partition should not (i.e. if two cells overlap then one is a subcell of the other). Mereology on the other hand makes no distinctions in situations where an object consists of parts that are not all connected (as Italy and its islands) or when an object is entirely inside or surrounding another object (as The Vatican and Italy); therefore it needs to be complemented with topological notions. In addition, if we quantify over wholes in a standard mereological framework, then we thereby quantify over all the parts, known and unknown, relevant and irrelevant, of such wholes. Granular partitions theory (Bittner & Smith 2001a) avoids these disadvantages via the intermediate formal machinery of cells, which adds to mereology the features of selectivity and granularity as well as topology, and precludes overlapping of cells that are not linked by the inclusion relation.

Granular partitions are ways of structuring reality (by dividing it up into meaningful chunks) in order to make it more easily graspable. The adjective granular refers to the possibility of identifying objects without having to recognize all their constituent parts. Granular partitions theory puts these cognitive devices and the objects to which they refer in two separate realms. Therefore it has two parts: A) a theory of the relations between cells and the partitions in which they are housed; and B) a theory of the relations between cells and objects in reality. Theory A defines a series of master conditions that all granular partitions must fulfil, characterizing them as rooted graphs without cycles and without upward (in the direction from leaves to root) bifurcations.

Theory B states that a granular partition recognizes an object *o* if and only if it has some cell *z* where the object fits, that is, if and only if z exists and *o* is the referent of *z*. Recognition presupposes two mutual directions of fit: from mind to world (projection) and from world to mind (location), i.e. if a cell z projects onto an object *o*, then *o* is located at *z*. Proper granular partitions are not ambiguous, that is, (at each level of granularity/abstraction) each cell can only project onto one (type of) object[1], neither (horizontally) redundant, that is, each object can only be located at one cell (a partition of celestial bodies including two cells –one labelled 'The morning star' and the other 'The evening star'- to refer to planet Venus would be confusing). Also their mereological structure (the way subcells are nested into cells) must

[1] 'Object' here is used in a wide sense, to include also scattered mereological sums. Thus a partition of the animal kingdom might include a cell labelled *man*, which projects onto that single object which is the mereological sum of all live humans.

reflect truthfully (although not completely) the mereological relationships on the side of the objects they recognise. Further details on the theory can be found in (Bittner & Smith 2001a).

Landcover maps in the form of polygon vector layers involve the construction of two reciprocally dependent granular partitions: a geographic partonomy, or *zonation*[1], defined over the mapped territory, and a partition of the attribute domain, or *taxonomy*, given by the map legend. The intimate relationship between classification and mapping, i.e. the fact that each individual scheme applied to the same landscape will yield a somehow unique zonation, was first recognized by Kuchler (1967a) in his seminal work on vegetation mapping, and is now widely acknowledged (Sinton 1979;Frank, Volta, & Mcgranaghan 1997). Using the notions introduced in this subsection, we are now able to specify some properties of both granular partitions (Bittner & Smith 2001a).

The zonation is complete, in the sense that there are no empty cells (every cell must project onto a portion of the mapped territory). The taxonomy on the other hand may not be complete, for it may possess some minimal cell that cannot be projected because that particular territory lacks the landcover type to which it refers. The reciprocal assertion, exhaustiveness, i.e. that every portion of the territory must be located at some cell, holds for both partitions, as long as i) the term *portion* has been clearly defined for the minimal cells of the zonation, implying that some *grain* requirements (a minimal size and topological connectedness and simplicity) must be fulfilled for portions to be recognised (more on this in 2.2.25); ii) there is a cell in the taxonomy labelled '*transition zone*' for recognisable (by the zonation) portions that do not fit in any of the cells of the taxonomy because they have a non-recognised mixture of categories; and iii) the case of selective maps (focusing only on e.g. seminatural vegetation or agriculture) there is a cell labelled *other* (e.g. non-forest, non-agriculture) that projects onto the zones of no interest.

Also, if the classification scheme is hierarchical, every minimal cell in the zonation has a corresponding minimal cell in the taxonomy. Note that in this case, adjoining minimal cells of the zonation may be aggregated hierarchically into bigger cells, that is, as we move from leaves to root in the taxonomy, the corresponding partition of the territory is potentially

[1] The term *zonation* (an arrangement or formation in zones), albeit may lead to confusion because its usage in Ecology (the distribution of organisms in biogeographic zones), is more specific than partition (a decomposition of a set into a family of disjoint sets), since zones are regions distinguished from adjacent parts by a distinctive characteristic.

simpler (with less zones). However, the zonation of a hierarchical taxonomy may not fulfil the (vertical) redundancy constraint, i.e. that a parent cell should not have only one daughter. Actually this will happen every time a region of the zonation has no neighbour with the same parent cell in the taxonomy. This problem can be tackled by letting only the most specific cell (the one farthest away from the root) project onto the region in question until it can be aggregated.

Finally, note that the representation of the zonation (by e.g. displaying the vector layer on a screen) preserves not only the topological relations of objects on the ground (due to the functional nature of the geodetic transformation used to map Earth surface features into a planar projection), but their mereological structure (the latter obviously holds only when the taxonomy is hierarchical) up to the minimal cells. Such perfection is due to the fact that the objects recognised are *fiat* (artificially delimited) objects carved out by the projecting partitions themselves. This fact raises doubts on the ontological status (their very existence in the real world) of the geographic objects onto which the cells of the zonation projects. These doubts are solved in the next subsection.

2.2.7. Geographic objects

In this thesis, the term *object* generally refers to a discrete spatial entity that has many permanent properties which endow it with an enduring identity and which differ in some way or another from the properties of its surroundings. **Geographic objects are complex** (having constituent objects-parts), **extended** (wider than high) **and of a certain minimum scale** (as to allow representation in a map) **objects on or near the surface of the Earth**, like cities, forests, lakes, mountains, agricultural fields, vegetation patches, etc (Smith & Mark 1998). In virtue of the commonsensical realist assumption of 2.2.2, we take for granted their existence, but there are at least two issues that may promote some doubts on the validity of this premise. The first one is raised by the fact that many of these objects have boundaries that are difficult to delimitate, therefore when we try to identify the region of geographic space occupied by those objects, we introduce some degree of arbitrariness that may lead to certain mind-dependence on their side. Second, the intuition that a volume of space cannot be filled by two or more objects simultaneously is in apparent contradiction with the proposition that two geographic objects may occupy the same region of a territory and yet may not be identical.

Before going on, it is worth noting some important points. **First**, the relationship between the region in which a geographic object is located (e.g. a forested area) and the object itself (the forest) is not one of identity. The forest can change its shape, shrink or even disappear, but a geographic region necessarily has the shape and size it has (Casati, Smith, & Varzi 1998). That is to say that a geographic object is a geographic region that can be identified for a certain period of time as the referent of a geographic name (e.g. the Sherwood Forest). As the territory evolves, the region onto which that name projects may change. **Second**, from the cognitive perspective, it is far more economical thinking in terms of separate wholes (objects) with distinct properties rather than in sets of connected points or cells whose content has to be determined cell by cell (Frank 2001f). **Third**, geographic objects may be nested, i.e. compounded of smaller geographic objects. However such decomposability is limited by size, since depending on the level of generalization applied to the territory, objects below a certain extension cannot qualify as geographic objects even if they are attached to the Earth surface. And **last**, geographic objects may have 'holes', i.e. parts that do not conform to the concept giving meaning to the object as an integrated whole.

As pointed out by Mark, Smith and Tversky (1999a), the fact that geographic objects are immovable (although they may change in size and shape because of losing or gaining parts) makes them inherit from space some of their ontological properties. In particular, location is one of the traits that distinguish a geographic object from all the others, marking its identity. This is in contrast to manipulable objects, which usually have location, orientation and even size as merely accidental. Another interesting hypothesis from these authors is that, again unlike manipulable objects, the immediate exterior of a geographic object may be significant for classification purposes (e.g. a forest surrounded by buildings may have the label 'urban park'). I would add that a) this dependence is inversely proportional to the size of the object (the former forest, should it have tens ok km, it would be no longer a urban park but a forest of its own), and b) not only the exterior but interior parts from a different category –later termed 'gaps'- may exert a significant influence on the final labelling of the object.

The shape, size and location of geographic objects is given by their external boundaries, which therefore contribute as much to their ontological make-up as do the constituents comprehended in their interiors (Smith & Mark 1998). **Many geographic objects are *fiat objects***, in the sense that they are delimited by boundaries which exists only in virtue of some human cognitive activity and that may not correspond to any observable discontinuity on the ground (Smith & Varzi 2000). Fiat objects may have some boundaries that are *bona fide* (genuine boundaries observable at a meaningful scale), but at least in some tracts they will have some fiat boundaries connecting the *bona fide* ones (e.g. the Mediterranean sea is a fiat object since the boundary separating it from the Atlantic ocean is a fiat line). Apart from the boundaries, fiat objects also exhibit a sort of fiat *continuity* (Smith 2001) in their interior parts, so that they are treated as if they were a homogeneous whole.

Some authors (e.g.(Rowe 1961)) argue that aggregates of objects like a forest are not objects in themselves but human constructs, and therefore should not be confused with integrated wholes like trees. However any given forest manifests both structure (patterns) and function (processes). The parts compounding its structure interact more strongly or more frequently between them than with the exterior. As such, the forest is an integrated whole, albeit such integration may be looser than the one of a single tree. Otherwise, the imposition of fiat continuity over the region occupied by the forest would be not only useless but deceptive, since it could lead to wrong inferences about the properties of that region.

Coming back to the first of the doubts that initiated this subsection, '**it is clearly true that the fact that an object is a fiat object does not entail that the object itself is mind-dependent, but only that some of its boundaries are**' (Thomasson 2001). Consider e.g. your hometown. The fact that it is administratively bounded by fiat lines does not mean that you are living like Jim Carey in the Truman's show. Moreover, many times such fiat lines lead to a stronger interaction between the constituents they bound and to a decrease of interaction with the constituents of adjacent units, think e.g. of the Muslim, Jewish and Christian quarters of Jerusalem. Also, geographic objects are supervenient on *bona fide* objects at lower levels, like trees or buildings, so that '**the interior of fiat objects are in this sense autonomous portions of autonomous reality**' (Smith 2001). In short, geographic objects are referents of geographic concepts. The fact that their boundaries can be drawn in many ways is more an epistemological problem than an ontological one.

The second doubt (how e.g. a geographic object 'forest *x*' occupies exactly the same region than the geographic object 'mountain y') can be easily solved using Frege's (1892) notion of sense and reference. A single portion of reality may be the referent of different names (recall the Venus example on the previous subsection), depending on the focus (e.g. landcover or physiography), which is given by the sense of the names. The fact that a thing can have different names does not preclude its existence in reality. But in order to insert such thing in a taxonomy, we can only use one sense at a time, otherwise we would be liable of double counting, and that is precisely what the theory of granular partitions prevents. The problem now is what region of space is exactly the referent of this name.

2.2.8. Sorites vagueness

Let us forget momentarily the forest and focus on the mountain, and let us imagine, adopting the example by Bittner and Smith (2001b), that the mountain is Mount Everest. It is not clear what parts along the foothills belong to it. One alternative is to hold that there are multiple candidates, all of them having the summit as part, where none of them can be said to stand as the legitimate referent of that name. The multiplicity of candidate geographic objects is a reflection of the vagueness of the name *Mount Everest.* In this thesis, Varzi's (2001c) view of vagueness as *de dicto* will be preferred to the *de re* view of fuzzy methods (Zadeh 1965), that is to say, vagueness will be treated not as a property of objects but as semantic property of language. Vagueness, rather than a defect of language, is an economic (it facilitates communication without cumbersome additions required to achieve precision) and epistemic (paraphrasing Heisenberg's principle, precision decreases the certainty of propositions) need. The reason for choosing the de *dicto* view is that the other (*de re*) requires further ontological commitments on the nature of fuzzy objects, complicating the mereotopological relations between them (Bittner and Smith 2001b). Besides, fuzzy set theory assumes that we can quantify membership functions (to what degree a given point does belong to an object) everywhere, an assumption that is not only is dubious, but leads to statements like e.g. 'this point of the Himalayas is 40% part of Mount Everest, 35% Mount Lhotse and 25% part of the valley', that are at odds with our entity view of geographical phenomena.

Figure 2-1. **Left: a partition, with cells *Everest*, *Lhotse* and *The Himalayas*. Right: A part of the Himalayas seen from space, with Mount Lhotse (left) and Mount Everest (right). From Bittner and Smith**

Turning back to the example, all admissible candidate regions must have the summit as a part and must form a unitary whole. Then a good algorithm to achieve such region would be: 1) locate the summit of Mount Everest; 2) initiate region *r* with only the summit; 3) scan the neighbourhood of *r*; and 4) if *x* is adjacent to *r*, then aggregate *x* to *r*. The problem is that there is no generally applicable stop condition that can be inferred from the concept *mountain*, and therefore we are confronted to what Brandon Bennet (2001d) coined *sorites*[1] *vagueness*.

2.2.9. Supervaluationism

A solution to this problem has been proposed long ago: *supervaluationism* (van Fraasen 1966). Sorites vagueness is characterised by the existence of border cases lying in the penumbra of a predicate, i.e. cases which are neither true nor false but indeterminate in truth value. Given a sorites predicate (e.g. 'John Smith is bald'), we can stipulate a crisping (*precisification*) of it (e.g. by sharpening the concept of bald person), i.e. a sharp boundary somewhere in the penumbra. From this we can decide that a predicate is unequivocally true, or supertrue (e.g. Yul Brynner is bald), if it is true for a set of representative precisifications, and the other way round with superfalse (e.g. Robert Redford is bald). In between, we can say that a predicate is in some sense true, if it is true for some precifisication (e.g. Bruce Willis is bald).

[1] Referring to the sorites (from *soros*, heap in greek) paradox, also known as little by little arguments. It was one of a series of puzzles attributed to Eubulides of Miletus (IV B.C.): Would you describe a single grain of wheat as a heap? No. And two? No. And three? No... You must admit the presence of a heap sooner or later, so where do you draw the line? (Source: The Stanford Encyclopaedia of Philosophy).

2.2.10. The egg-yolk representation of regions with vague boundaries

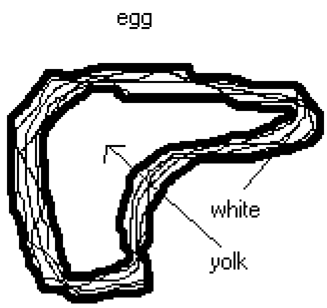


Figure 2-2. Egg-yolk representation

In the case of Mount Everest, each precisification (effected e.g. by a cartographer drawing the boundary on a map) would yield a zonation cell (defined in 2.2.6) that would project onto a crisp region of the Himalayas including the Everest summit. If we had a team of renowned cartographers working on the issue, one way to reach consensus would be to ask separately each one to draw his/her own version of the boundary. Then we could draw two wrapping lines encompassing all the precisifications. Both lines would yield two concentric regions alike a fried egg (that is why Cohn and Gotts (1996b) coined this kind of representation 'egg-yolk'). The egg is the maximal region that Mount Everest can occupy under all admissible precisifications. Then the predicate that 'a part of the Himalayas outside this egg is part of Mount Everest' is superfalse. Conversely, that a region inside the yolk is part of Mount Everest is supertrue. Then the border to be agreed will lie somewhere along the white of the egg.

2.2.11. The epsilon-band model of positional boundary error

Let us leave Mount Everest and go back to the forest. An egg-yolk representation could also be achieved in this case, by e.g. asking a team of photointerpreters to delineate the boundary of a given forest. In sectors where there is a clear discontinuity between the forest and the surroundings (e.g. an agricultural field), the width of the white will be quite narrow, whereas in sectors where the forest grades e.g. onto a shrubland, the different interpretations will lead to a wide white. Therefore the width of the white can be assimilated to a measure of the boundary positional error, i.e. to a *probabilistic epsilon band*[1] within which the 'true' boundary of an object has a probability of 100% of being located under the supervaluationist conception. An example of the use of epsilon bands in this way can be found in Green and Hartley (2000c). They calculated the width of epsilon bands by measuring the positional error introduced by georeferencing, digitising and subjective interpretation, and found that the latter process accounts for 90% of the total error.

[1] This concept was introduced by Honeycutt (1987) as an application of Perkal's (1966) work on cartographic generalisation to the estimation of boundary positional error.

2.2.12. Rough projection

Now imagine that instead of a single forest we have to produce a zonation (a granular partition projecting onto a portion of earth's surface) over a whole territory. In practice there is no feasible way (in a mapping project context) to determine the width of the epsilon band along all the boundaries, so we shall have to choose arbitrarily one of the precisifications (e.g. the one made by the most experienced interpreter) as the zonation to be imposed on the territory. In 2.2.6 we assumed that projection was an exact function, i.e. that the cells of the zonation project exactly onto the geographic objects to which they refer. But there are no such things as infinitely thin lines in geographic space.

Even considering, as assumed in 1.12, that geocoding and cartographic system of reference are perfect, the thin line drawn by the interpreter, when transferred to the 1:1 scale, will have a considerable thickness on the ground. Moreover, even if we had the line digitised on a screen so that the exact coordinates of the vertices are available, we would get again a thick line, of width equal to double the accuracy of the positioning instrument, ranging from the 20 m of a standard GPS to less than 0.5 m for some DGPS[1]. That is to say that, even assuming the existence of geographic objects that are the counterparts of crisp fiat objects created in the image by photointerpretation or quantitative analysis, the projection of these crisp objects onto the ground is rough. Consequently, there will be always an epsilon band bounding the geographic objects carved out by the partition, which will stand for as an approximation of the boundaries of the zonation cells. As a corollary, note that whenever a geographic point is mentioned in this thesis, it will actually refer to a ground circle of some 0.5 m radius centred at the point coordinates.

[1] Differential Global Positioning System. It consists in keeping stationary a GPS receiver at a fixed known location. Individual range corrections for each satellite seen by the fixed GPS are then sent by some form of telemetry to the mobile GPS receivers and applied in real time to greatly increase the position accuracy of those mobile receivers.

2.2.13. Rough location

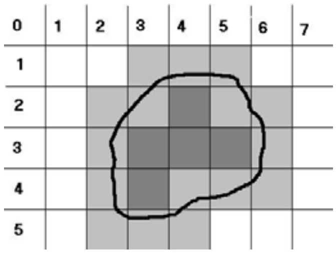


Figure 2-3. Rough location in a raster reference partition

Something similar occurs to location, due to the limited resolution of the zonation representation. To explain it, let us come back to the forest and suppose that its boundaries have been precisely (say with < 1 m accuracy) surveyed from the ground. Then imagine that we want to confront this ground-based zonation **`Zground`** (that for the moment will consist of a minimal cell, the forest, and a root cell, the territory on which the forest lies) with a hypothetical image-based zonation **`Zref`** that could be produced e.g. from a 250m resolution MODIS image. The image can be seen as a regular raster-shaped zonation in which each cell is a pixel that is projected onto a square plot of terrain of *p* meters side, 250 m in the example. In order to compare both zonations, they have to be represented at the same scale and cartographic projection (e.g. 1:25000, UTM). With this view we can use the blank image raster as the (Cartesian) frame of reference (i.e. a grid template) within which the forest is to be located (fig 2-3).

Rough location[1] describes the location of the geographic object 'forest', previously delimited by the ground-based zonation, within the set of square terrain plots defined by a raster zonation. In order to proceed, three mutually exclusive primitive relations between cells of both zonations have to be defined (Bittner & Winter 1999): *full overlap* (***fo***, when a cell *z* –pixel- of the reference partition is completely included in the ground-based zonation cell –forest), *partial overlap* (***po***, when the pixel is crossed by the boundary of the forest), and *no overlap* (***no***, when the other two do not hold). In figure 2-3, these situations correspond respectively to dark grey, light grey and white pixels. With these primitives, which can be directly expressed by a binary function ($f(z,o)=1$, if they hold, 0 otherwise), the rough location of a geographic object o can be represented by upper ($L^U(o)$) and lower ($L_L(o)$) approximation sets as follows:

$$L^U(o) = \{ z \in \mathtt{Zref} \mid po(z,o)=1 \vee fo(z,o)=1 \} \qquad (2.2.13.1)$$

$$L_L(o) = \{ z \in \mathtt{Zref} \mid fo(z,o)=1 \} \qquad (2.2.13.2)$$

[1] Introduced by Bittner (1999e) as an extension of Pawlak's (1982c) rough set theory.

In this thesis, the lower approximation will be preferred, for two reasons. First, the certainty that all the parts (pixels) included in the representation of object *o* actually belong to it, is obviously higher than in $L^U(o)$. Second, by letting out partially overlapping pixels, objects will be bounded by a 1-pixel-wide stripe that will be treated in this thesis as an epsilon band of fixed width. The former analogy is valid since on the one hand, the 'real' (from Zground) boundary lies somewhere within those pixels, and on the other, no reliable decision can be made a priori about the membership of these pixels to one of the adjacent objects they separate (recall the mixed pixel problem of image classification).

2.2.14. Rough zonation

Let us now extend the ground survey to the whole territory, so that every geographic object of interest is foregrounded on *Zground*. This zonation now is a complete partition consisting of contiguous irregularly shaped crisp cells, each one projecting onto a definite region of that territory. We can use again the image grid *Zref* to roughly represent this partition:

$$K(Zground) = \{ z \in Zref \mid (\exists\, o \in Zground \mid po(z,o)=1) \} \quad (2.2.14.1)$$

Where K(*Zground*) is the rough morphology of *Zground* as represented in *Zref*, i.e. a raster mask where pixels containing an object boundary are set to 1, and 0 otherwise. If each object *o* from *Zground* has been identified with an unique numeric label j (so that there is a labelling function *l*(*o*)=j with j=1,...k , being k the total number of objects/regions), then a rough replica, or *rough zonation*, of *Zground* can be obtained from the morphology, or directly from the overlap functions, through the following mapping:

$$W_p : Z_{ref} \times Z_{ground} \rightarrow \aleph \;\; ; \;\; W_p(z,o) = \begin{cases} 0 & \text{if } \mathrm{po}(z,o) = 1 \\ \\ \ell(\mathrm{o}) & \text{if } \mathrm{fo}(z,o) = 1 \end{cases} \quad (2.2.14.2)$$

The result of applying W_p to *Zref* is a raster in which each cell (pixel) has as value either the numeric label of the object fully overlapping with it, or the null value, if it overlaps with more than one object. The suffix p of W_p refers to the width of the side of the terrain square plots, i.e. the pixel size. In chapter 3 W_p will be assimilated to the watershed partition.

2.2.15. Junctions and arcs

To end up the 'rough' discourse, two more definitions will be added. A *junction* is a pixel from W_p where three or more objects meet, i.e. a 0-valued pixel that has in its 4-neighbourhood (not including the diagonals) more than two 0-valued pixels. An *arc* consequently is defined as a chain of 0-valued pixels bounded by two junctions. These concepts will be used in Chapter 3 to derive a vector layer from W_p.

2.2.16. Class attributes

In the following subsections, we will investigate what criteria may be used to delimitate each fiat object, and how the drawing of boundaries can be performed. Recall the forest example. Before putting the GPS to work, the surveyor must have a clear idea of what a forest is: how dense (% cover) and how high the vegetation must be, from which species, and how large the area populated with trees must be in order to consider it for inclusion in the map. So the vague natural language concept of *forest* must be sharpened as to avoid indeterminacy of boundaries, and this is done modulo the classification scheme used to group different landcover configurations.

In order for the scheme to constitute a proper granular partition, i.e. to avoid ambiguity and redundancy, **each class must differ from the others in at least one *measurable property*, or *attribute*, and more importantly, the classes must be defined in such a way that can be distinguished with the means used in the compilation of the map, so that the classification criteria can be converted into measurable attributes. The latter should be not only easily observed, but of maximum significance for the model of landscape we want to convey**. After choosing the set of attributes, patches conforming to the definition of a given class may be allocated to the mereological sum of that class, once the proposition 'patch *x* complies with the admissible range of values for class *i* in all attributes' has been verified.

2.2.17. Class definitions

Landcover class definitions are supervaluationist precisifications of vague concepts from natural language. These concepts may include just a name, as e.g. *forest*, or a compound proposition, as e.g. *broad-leaved temperate forest*. The sharpening is made by defining a series of relevant attributes and set of admissible values of each one of them for each class. The expected result of the sharpening is a partition of the geographic space into crisp objects that can be taken as instances of the defined classes. I will explain how this is done with a clarifying example by Brandon Bennet (2001d). To put the example into context, imagine that we are compiling a landcover map over some territory of some 100 km^2 extension, which for simplicity reasons is flat. The map has only two classes, forest and non-forest, although it will be produced from a comprehensive ground survey.

Let us assume, holding again a commonsensical realist view, that the relevant measurable properties of forests are supervenient on the relative abundance, spatial distribution and measurable properties of some individual entities called trees. Therefore the first step towards a sharpening of forest is a precisification of the concept 'tree', e.g. 'a woody perennial plant having an elongate main stem and total height greater than 5 m'. Then we can formulate a precise definition of the object class 'forest' with the following two statements: i) *a forested cell is a topologically convex area of certain minimum extension that is densely covered by trees*; and ii) *a forest is a geographic object whose terrestrial extension is a maximal set of connected forested cells*. The 'maximal' adjective of second statement precludes overlapping between forest objects, i.e. a forest cannot include parts that can be considered as a forest on their own.

2.2.18. Measurement disks

The 'certain minimum extension' of the forested cell must be such that the 'density' attribute can be sensibly measured in it. The *minimum minimorum* would thus be double the mean spacing between trees (see why in 2.2.23). The *maximum minimorum* would be such that the positional accuracy of the future boundaries is not decreased significantly (2.2.21). Then the actual size has to be selected between both extremes, so that the resulting measurements are homogeneous over a zone that is considered to be homogeneous (2.2.22). Another decision that has to be made is the shape of the convex area where the density has to be measured (circle, square, rectangle, hexagon, etc). Suppose we set it to a circle of radius 17.84 m, equivalent to an area of 0.1 hectare. In general, unlike other geographic attributes like altitude or temperature, **the attributes taken into account in landcover classification schemes cannot be measured at a given geographic point without inspecting a considerable larger area around the point** (Goodchild 1994). **This area of observation will be defined**, adapting Bruegger's (1994a) notion of resolution disks (appendix 3), **as a circular template of diameter *d* centred at each measured point, hereafter called *measurement disk***. The size of the disk is attribute dependent, and it is closely related to the classification disk introduced in chapter 1, which is equal to the diameter of disks used to measure the attribute requiring the largest area of observation. Finally, note that measurement disks are a special kind of *support*[1] that is invariant to the orientation of the data, and have the same circular shape than the support (i.e. the GIFOV) of most remote sensors.

2.2.19. Attribute measurement

Coming back to the forest example, once the measurement disk size is defined, we can formulate a mensuration protocol for the density attribute (hereafter the *tree cover fraction*, TCF) at any given point of the geographic space, as follows: a) establish a circular plot of 17.84 m radius (1000 m^2) centred at the point; b) measure the top height of all the plants inside the plot potentially qualifying for 'trees', c) select those candidates measuring more than 5 m; d) project orthogonally onto the ground the outermost (in respect to the stem) parts of the crown of the selected candidates; e) measure in m^2 the ground area covered by the interior parts of this projection; f) divide by 10 (so that the fraction is expressed in %). Other attributes could also be measured in one way or another within this measurement disk.

[1] The term *support* is used in Geostatistics (Matheron 1971) to refer to the geometrical size, shape and orientation of the regions from which the measurements are drawn.

2.2.20. Geographic fields

A geographic field is a regionalised continuous variable representing some observable property of a territory, which has been unambiguously defined as to allow precise measurement at any geographic point. Geographic fields are in general ontologically dependent on the (scattered) distribution within the territory of some discrete objects (as e.g. trees), although in some special cases they can be derived directly from continuously distributed physical magnitudes (as e.g. surface temperature).

Now imagine we have unlimited budget and manpower to compile the map of the example. In this scenario, we could calculate, using the specified measurement disks, the tree cover fraction at all the points of the territory, and the output of this activity would be an exact version of the TCF field.

In a slightly less optimistic scenario (a huge but finite budget), we could limit measurement to only the centre of the cells of a square grid superimposed on the territory, choosing a sampling interval such that no point is left untouched by the measurement disks (25.23 m for a 17.84 m radius). The value at non-sampled points could be estimated by some regional interpolation (kriging) technique.

This version of the field would be practically identical to the former, but as we extend the sampling interval, the resulting versions may be increasingly inconsistent with the exact field, up to a point (e.g. when the sampling interval is greater than the mean size of forests) where no information at all can be derived about the boundaries or even the existence of small forests.

Note that the extended nature of measurement disks guarantees the mathematical differentiability of the field over its entire domain, that is to say, even having a forest with sharp boundaries on the ground, the transition to zero will be gradual along a width equal to the diameter of the disk.

2.2.21. Object demarcation via field thresholding

Once the field is constructed, we only need to precisify the predicate 'densely covered' by setting a threshold on TCF. Imagine for now that the first object we will delimitate is a 100 ha plantation forest with trees regularly distributed and completely covering the ground. Each density precisification can be represented by an isoline on the TCF field. Each isoline consists of a set of connected points having as TCF value the chosen threshold. Then the extent of the plantation forest would be the area enclosed by this line. In other words, what we have done is to derive a dichotomous (with only two values, 1, for points above the threshold –forest, and 0, for points below it –non-forest) field, or ***classified field***, from the original one. What the classified field actually does is to foreground the objects of interest.

In order to evaluate the positional accuracy of the threshold-derived boundary, let us assume that the 'true' boundary is the convex hull (involving polygon) of the ground projection of the outermost (with respect to the forest) parts of the crowns of the trees standing on the forest edge. If we are extremely exigent and set a minimum cover fraction (mTCF) of 100%, the resulting boundary will be displaced some 17 m towards the interior of the true boundary. Conversely, if we consider forest anything where there are trees and set mTCF to simply > 0%, the displacement would we in this case 17 m outwards. The boundary obtained by other precisifications will lie in between both extremes. If we set e.g. the threshold to >50%, the boundary will coincide with the true edge, but if the plantation has a TCF of 50%, we will return to the previous situation. Therefore we can conclude that the positional accuracy of the threshold-derived boundary will depend on i) the diameter of the measurement disk, ii) the threshold value chosen for the attribute, and iii) the actual value of the attribute within the object. Note in any case that the maximum positional error, when the measurement is complete (made at every point), is equal to the radius of the disk, provided the value of the attribute is uniform throughout the object.

2.2.22. Spatial homogeneity

Before examining the more general case where the trees are unevenly distributed, the concept of homogeneity should be clearly stated. Homogeneous means 'of uniform structure or composition throughout', where 'uniform' stands for 'not varying'. It is important to note that

homogeneity is an observational (epistemic) property rather than an inherent (ontological) property of objects. That is to say that the homogeneity of an object is dependent on i) the scale of observation; and ii) the way the observer picks out regularities and traces over differences within the structural components of the object. The first condition implies that every object may appear homogeneous (smooth) if it is observed from a sufficiently long distance. As the object is approached, it will begin to manifest texture until its constituents become resolved and hence its varying structure appreciated. Reached this point, it is the observer who imposes fiat homogeneity by foregrounding the parts that are evenly distributed and tracing over parts that are not. In this sense, all the geographic objects that fit into the minimal cells of a zonation are homogeneous by definition, since no further parts are recognised.

Depending on the focus, various judgments about the homogeneity of objects can be made. Each judgment will generally refer to a single measurable property (attribute) of objects (e.g. cover, dominant height, species composition). The homogeneity of an attribute over the region of space occupied by an object has two aspects: the magnitude of the variation and the spatial distribution of the variation. The first one can be estimated through the coefficient of variation (CV, the standard deviation divided by the mean) of a systematic (regularly distributed throughout the region) sample of measurements. The sampling interval should be such that there are a *sufficient* number of measurements. The measurement should be defined so that if the property measured by the attribute is considered by the observer as homogenously distributed within the region, the CV of the sample is *low enough*, and the other way round when is considered inhomogeneous. Obviously, the predicates 'sufficient' and 'low enough' have to be specified by the observer.

The spatial distribution of the variation can be evaluated through a variogram (Matheron 1971). Variograms are plots of the semivariance of a spatial data set against distance. Semivariance is defined as half the mean of the squared difference in value of an attribute z, between N(h) pairs of points i and j separated by distance h (usually called *lag*). It typically increases with increasing distance, and is inversely related to spatial autocorrelation. The variogram is regarded to 'provide a concise and unbiased description of the scale and pattern of variability in a spatial data set' (Curran 1988).

2.2.23. Admissible disk size

Let us come back to the example. The main factor contributing to the value of CV, apart from the spatial distribution of subobjects (trees) within the region, is the size of the measurement disks. The TCF of a plantation forest will always be considered as homogeneous, independently of its actual value, therefore the corresponding CV should be low enough (say < 5%). Suppose that a) we use a 1m sampling interval, b) the plantation frame of the forest is of one tree each 5 m, and c) the mean crown diameter is 2 m, so that TCF is 50%. It can be shown that whenever the disk radius is greater than 5 m, the resulting measurement will approximate TCF=50% for most points, with CV approaching asymptotically zero as the disk size is enlarged[1]. As we reduce the disk size, CV will be increasingly higher[2]. So the lower admissible bound for a disk diameter in order to correctly measure TCF is equal to double the mean spacing between trees.

Regarding the upper bound, as the size of the disk is enlarged, the percentage of sampling positions where part of the disk is located out of the forest will increase. Therefore, if the forest is surrounded of e.g. agricultural fields, the mean TCF will be underestimated. Besides, as stated in 2.2.20, the positional error of boundary placement will increase. Consequently, the upper bound could be set to the maximum mean spacing between trees that can be found in the most sparse of the forested regions eligible as a forest. To avoid further complications and arbitrariness, I will conclude with a simpler statement, adopting the general rule-of-thumb used in forestry to obtain a sufficient sample size: **the disk diameter should be big enough to capture at least 20 trees**. This means that for most situations, disks of size between 500 m^2 (25m diameter) and 2000 m2 (50m) will be suitable for forest mapping, and the positional accuracy obtained would be similar to e.g. that of a georefererenced Landsat image.

In the example we had only one attribute, TCF, measuring a property of landcover that is supervenient on the properties (crown size) and spatial distribution of individual trees. The disk size is chosen so that measurements are homogeneous over a zone that is considered to be homogeneous with respect to the property measured by the attribute. In general, the classification will be based not on one but several attributes. Each one will have an optimum

[1] This is just an approximation, for in order to include in the measurement the crown of a tree, its stem should be located within the disk, so that TCF will systematically underestimated until the disk has a radius several times greater than the tree spacing.
[2] Note again that the underestimation of TCF will increase steadily with a reduction of the disk size, the value of TCF approaching asymptotically zero.

disk size in relation to the former assertion. If the optimum size of all the attributes is in the same narrow range, a unique disk size can be used to construct all the fields involved. Otherwise, each field would be preferably constructed using its optimum disk size.

2.2.24. Gaps and islands

Now we are ready to address the case where there are natural forests in the territory with trees unevenly distributed. Within them, there may be pockets, or *gaps*, where there are no trees at all. Even in the plantation forest there may be gaps consisting of clear-cuts. Consequently the TCF threshold used to precisify the forest may produce more than one isoline per target object. For each forest there will be an outer isoline, defining the exterior boundary of the forest, but there may be one to many inner isolines circumscribing possible gaps. In other words, the objects of the classified field may have holes. The concept of gap can be generalised by saying that **a gap is any interior part of a geographic object, lacking enough extension as to conform a geographic object of its own, where the value of the relevant geographic field(s) is out of the range of admissible values for objects of this class**.

If a sparse woodland continues the forest of the example and we move towards it, the abundance and size of gaps will increase. Once in it, if trees in the woodland are grouped into coppices, gaps may become connected. In this case, the area enclosed by the isoline is better viewed as an island than as a gap. Islands exhibit the same problems than gaps, and they could be defined in the same way. They could even be considered gaps, where the bigger geographic object in which they are located, in the absence of an ampler involving isoline, is the root cell of the zonation, that is, the extension enclosed by the border of the map. In this case, the bigger region in which the objects of interest are littered would constitute what ecologists call the *matrix*, or background cover. In general, gaps and islands are holes in the objects foregrounded by a classified field. Finally, note that very small gaps or islands (say smaller than the measurement disk) will be absorbed by the disks, therefore they will not qualify for the former definition. The problem now is to decide whether the emerged gaps or islands deserve representation.

2.2.25. Minimum mapping unit (MMU)

Gaps and islands are subject to sorites vagueness, therefore they need a precisification regarding its minimum size. To see the problem, let us look to the photographic negative of the former example. Now we have a geographic object such as a big farm and we have a woodlot forming an island in it. The question is how large the woodlot must be in order to qualify as a geographic object in our forest/non-forest map. The answer relies on the *minimum mapping unit* (MMU), a cartographer's choice indicating the level of generalisation of the map. **MMU is the minimum size** (or sometimes width, when referred to elongated objects) **that an object** (represented in the map by a polygon, or mapping unit) **must have in order to get into the map**. By imposing a MMU size, the mapmaker is saying that any region below this size does not qualify as an instance of the geographic objects included in the legend of his/her map. Three factors are involved in this choice: the goal of the map, the available budget for the mapping project and the spatial configuration of the territory.

The goal of the map (e.g. land planning, agriculture, forestry, wildlife conservation) determines the nature of the objects of interest (e.g. landcover patches, agricultural parcels, timber stands, vegetation patches) and the level of detail with which they are to be studied. The budget constrains this level of detail, limiting e.g. the number of sites that can be included in the field survey, and the map scale (if e.g. it will be printed for distribution). Finally, **the MMU has to be compatible with the size and distribution of the objects of interest within the territory**. For example, suppose that the map is conceived as a means for allocating surfaces to a network of timber measurement plots (as it is actually done in many national forest inventories). Additionally, grant the MMU 2 ha and the woodlot 1 ha, so that it is not eligible for representation. If this situation (a small woodlot within a farm) were rare in the territory, the dismissal of small woodlots would not affect the overall result of the inventory. But if the region were full of small isolated woodlots that are regularly exploited for timber, the selection of a MMU bigger than the mean size of the small woodlots would lead to a severe underestimation of the timber stock of that region.

After balancing these three factors, a final choice for MMU is made. Common MMU size ranges from 0.5 ha (e.g. the US National Park Service Vegetation Mapping Program) to 25 ha (e.g. the EU CORINE Land Cover Project). Note that several class-dependent MMUs can be defined for the same map. In this case, classes of special interest (e.g. high diversity habitats), distributed in small patches, are assigned a smaller MMU. Nevertheless, the general rule is a single MMU. If the map is to be printed in a paper series, MMU has to be big enough as to be representable at the scale of the series (say > 10 mm^2). Another cartographic limitation for polygons close to MMU is a smooth, preferably convex shape. Note that, since GIS enable visualisation at any scale, the only clue, in the absence of metadata, about the intended scale of a vector layer, is the area of the smallest polygons, i.e. the MMU (and perhaps the mean interval between vertices). In the context of GIS, MMU is more a cognitive need than a representational constraint, avoiding excess of detail and subsequent confusion.

In any case, MMU is one of the main factors affecting the information portrayed in thematic maps. **Each MMU size will lead to a different model of the territory**. If the model is used e.g. in a landscape ecological study, the conclusions drawn will depend not on the territory but on the model, therefore they can vary significantly with different MMUs. As a point of fact, Saura (2001) analysed the influence of MMU in several commonly used landscape indices (mean patch size, edge length, inner edge density, perimeter-area fractal dimension, etc), and found that most of them were highly sensitive to variations in MMU. Also, Stohlgren and Chong (1997), mapping the vegetation of a study area in Rocky Mountain National Park, USA, noted that, as the size of MMU increased, the estimated number of plant species and habitat patches (polygons) decreased. These variable results are related to the Modifiable Areal Unit Problem (MAUP).

2.2.26. The Modifiable Areal Unit Problem (MAUP)

We saw that, on the one hand, different disk sizes and sampling intervals produce diverging fields, and on the other, different zonation methods yield differing partitions departing from the same fields. This lack of unique solution is known as the Modifiable Areal Unit Problem, or MAUP (Openshaw 1984). MAUP illustrates the sensitivity of analytical results to the definition of the areal units from which data are collected. It arises from the fact that these units are arbitrarily defined and eventually modified to form larger units. Therefore, if the areal units are arbitrary and modifiable, then the value of any study based upon them may be rightly questioned. MAUP was identified in the context of socio-economical geography, but is has been also found in landscape ecology and remote sensing (Jelinski & Wu 1996). Marceau (1999) gives a comprehensive review on the issue.

In geographic fields, MAUP is manifested by the different configurations that the field may adopt depending on the disk size and the interval between sampling positions. Enlarging the disk size (analogue to smoothing an image) creates a new field. The problem of how the spatial variation in the new field relates to that of the original one is a special case of MAUP known as COSP (Change Of Support Problem) in geostatistics (Matheron 1983;Cressie 1996). In zonations, MAUP is evidenced by the multiple possibilities on defining the boundaries of the zones. Similarly, in remote sensing, MAUP shows up both in the processing and the analysis of the images. In the processing, it appears in relation to a) pixel size selected for a particular study, and, when this size differs from the one of the imagery, b) the resampling method applied in order to have the imagery resized to that pixel size. In the analysis, MAUP is inseparable of a) the segmentation method chosen and b) the stop criterion relative to the size of the segments (given by e.g. a minimum and/or maximum size or by a fixed total number of segments). Given a target size of final segments, each method will yield a somehow different partition departing from the same set of images.

Regarding the solutions proposed to mitigate the MAUP effects, Openshaw (1984), rejecting the premise of objectivity in the design of zoning systems, proposed a method that starts by formulating an hypothesis concerning the expected result for a given model, and then aggregating areal units to the point where the target result is attained. Under this approach, the definition of an optimal zoning system changes with the kind of problems under investigation. Other solutions were proposed by Fotheringham (1989b) and Visvalingam (1991), who respectively suggested the identification of basic geographical entities, and the use of basic spatial units that define the spatial primitives of the phenomenon under study. The approach followed in this thesis is in the line of the suggestions of these three authors.

Another recommendation is to systematically perform a sensitivity analysis in order to provide the range of results obtained when different areal units are used. However, the number of variables, scales, and zoning alternatives is usually overwhelming, making unfeasible such analysis. Notwithstanding it, at least it should be carried out on a limited set of cases, so that the validity of the conclusions can be appraised properly. In some applications, showing only the results originating from the use of one set of areal units can hide the most significant aspects of the problem and cast doubts on the conclusions that are presented. From Marceau's (1999) review it seems that, contrary to the pessimistic opinion that MAUP effects are intractable (Wong & Amrhein 1996), it is possible to control and predict these effects to some extent.

2.2.27. Mosaics, conglomerates and facets

Once the MMU is defined, all the gaps or islands smaller than this size will not qualify as geographic objects and therefore will not be represented on the map. Consequently, information regarding the presence of alien regions within objects will be lost, as a parsimonious exchange for clarity. The trade-off is valid as long as the total area occupied by gaps within the object is low enough, otherwise the loss has to be compensated with some addendum to the object's label, reflecting a significant presence of gaps. In extreme cases, it is the content of gaps/islands what marks the final label of the object (e.g. 'woodlot conglomerate', where the farmland is the 'matrix'). **These heterogeneous geographic objects are usually called *mosaics*, where heterogeneous means that their areal percentage of gaps is above a certain threshold**. Within a mosaic, there is a lack of information regarding the actual spatial distribution of gaps, but in turn their general pattern and areal percentage can be conveyed through the label.

Notwithstanding the foregoing, the term 'mosaic' is better fitted to the case where there are a complex of contiguous geographic objects of different type and of size close to MMU. The common solution to this situation is to amalgamate the small objects into a bigger heterogeneous object with a compound label reflecting the nature and distribution of the mixture. Note that an object containing a (recognised) mixture of subobjects (e.g. birch-spruce mixed forest) may not constitute a mosaic, providing the blending takes place in regions smaller than MMU that are uniformly distributed across the object. The presence of mosaics in a map depends not only on the spatial configuration of the landscape but also on the level of detail of the associated taxonomy (legend) and the chosen MMU. In general, for given level of taxonomic detail, **the larger the MMU size, the greater the fraction of the territory catalogued as *mosaic***.

In order to distinguish between both types of mosaics, the first one could be called *conglomerate*, although the term *mosaic* may be applied generically to both, whenever the distinction is irrelevant. Conglomerate (in geology, a rock composed of rounded fragments in a cement) captures better the existence of a 'matrix' in the first case. Finally, another term will be added to refer to each one of the pieces that compound a mosaic: the *facet*. A mosaic has usually many facets, but they are only from a few types, at least two. The term facet applies better to the second case of mosaics described in the former paragraph, but in the case of conglomerates, the gaps/islands can be equated to facets.

2.2.28. Geographic models and the first law of Geography

Having set forth the conceptual basis for the idealistic version of the model of geographic reality proposed later in this thesis, I will introduce now the main points underpinning the realistic version. A model is a formal representation abstracted from a piece of reality, or as it was defined in appendix 2, a model is communicable structural information. If that piece of reality is a territory, then we are talking about geographic models. **Geographic models answer three basic questions about the territory**: **what** things or facts there are in it, **where** they are and **when** this *statu quo* took place. These questions correspond to the three interrelated components of geographic information identified by Sinton (Sinton 1979): **theme**, **location** and **time**. Sinton established a general set of rules to construct geographic models based upon those components. In Sinton's view, measuring one of them requires that variation within a second component is systematically controlled, and the third one is fixed, or in a sense, ignored. While Sinton's concept of 'controlling' a component is (at least to me) little intuitive and even bizarre, an important point of his analysis is that it stresses the dependence of themes on space and time. As Barry Smith (1995b) puts it, 'There is, in other words, a relation of foundation or existential dependence between sensible qualities and spatio-temporal extension (no colour can, as a matter of necessity, exist without spatial extension, no sound without duration, etc)'.

Time is normally fixed in maps because they usually are snapshots of the territory taken when the data (e.g. aerial photographs) were collected. According to Sinton (1979), thematic maps 'measure' the location of the objects of interest, given a systematic control of how to define them categorically. In this case, time is fixed and theme is 'controlled', whereas location is being measured. In Remote Sensing images, time is fixed and location is "controlled" by imposing a raster grid over the territory, so that theme (e.g. albedo) can be measured. 'Theme' means here the same than 'sensible quality' in the previous quotation, or than 'geographic field' as defined previously. Therefore themes are potentially measurable throughout all space and time, even if the measurement values are zero or not available at some locations and times.

Location is usually modelled as a Euclidean two-dimensional space, where each dimension corresponds to the latitude and longitude given by some cartographic system of reference. With this planar geometry (that controls location) and a synchronic view (that ignores time), Sinton's themes, or geographic fields as defined here, become the primitives of any geographic model, since their measurement yield the structure upon which the model is constructed. At each location, a plethora of themes or attributes can be measured, each of them related to some particular aspect of geographic reality. Each attribute is a regionalized (distributed in the fore mentioned 2D space) variable that changes in a continuous manner from one location to the next. Continuity is guaranteed by the commonsensical observation that ***'everything is related to everything else, but near things are more related than distant things'***, which has come to be known as the ***First Law of Geography***[1].

On the other hand, **qualitative differences are supervenient upon quantitative ones**. In Barry Smith's (1995) words, '*the sensible qualities of objects can in every case be identified with the properties of certain corresponding physical variations'*. Hence, differences between two given adjacent geographic objects ultimately rely upon differences between (at least one of) the quantitative attributes that define the classes to which each one of them belong. Considering the foregoing, an additional law of naïve Geography could be stated as follows: ***boundaries drawn in maps are the places where the first law of Geography is violated***. Such places, hereafter called *singular points,* correspond to zones where the fields change abruptly at the relevant scale of observation, and hence can be interpreted also as qualitative discontinuities. The latter are explained mathematically by René Thom's (1975) theory of attractors (also known as catastrophe theory). Its application to geographic fields will be briefly outlined in the next subsection. The account is an adaptation of the more general one from Barry Smith (1995).

[1] Proposed by Tobler, Waldo, 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46(2), 234-240. Cited by Mark, D.M. and Egenhofer, M.J. 1996. Common-sense Geography: Foundations for intuitive Geographic Information Systems. GIS/LIS'96 paper (NG-I-21).

2.2.29. Object demarcation via Thom's morphology

The merit of catastrophe theory, whose application to semiotics is sketched in (Thom 1988), is that it provides the link between the quantitative physical theory of the world and our qualitative daily experience via which we construct our commonsensical view of the world. The following excerpt from B. Smith (1995) states the point better:

"... Physics, for all that it is restricted to the quantitative, does indeed deal with the phenomena from out of which the qualitative world is composed. What it does not do is to deal with those specific ways in which these phenomena are composed or knitted together or are delineated from each other that are relevant to the world of our qualitative experience. It is in showing how to fill this gap that Thom's achievement lies."

According to this theory, every object, or physical form, can be represented as an attractor of a dynamical system on a space of internal variables. Such an object is stable, and so can be recognized, only when the corresponding attractor is structurally stable. The stability is attained by a process, called ***morphogenesis***, consisting in the disappearance of the attractors representing the initial unstable forms, and their replacement (by capture) by the attractors representing the final form, which is the observable state of the object. A system compounded of several objects will have several stable attractors that together define the observable steady state of the system, manifested through some stable patterns. The attractor can be thought as the centroid of the object, so that points belonging to the object are more attracted to it than to the centroids of neighbouring objects. However there can be points in the boundaries between objects where this attractive force becomes unstable and a infinitesimal move in one or another direction may produce a change of attractor: these are the singular points.

Topologically speaking, the application of Thom's theory to geographic fields could be as follows. Let W be the planar representation of a territory in some cartographic system. The different qualities that can be observed in the territory are represented by n geographic fields q_i obtained from measurement disks, each a function $q_i(w)$ of points $w \in$ W. A point is called *regular* if all the $q_i(w)$ are continuous in a neighbourhood of w, of radius close to the size of disks. Let R be the set of regular points of W. R contains a neighbourhood of every one of its points, therefore it is an open subset of W. Let K be the complementary set of R relative to W. K is the closed set of non-regular points of W. By definition, w is a non-regular point if and only if there is at least one field q_i that is discontinuous at *w*. It can be easily shown that these non-regular points are the singular points mentioned in the previous paragraph.

To see it, imagine a territory consisting of a mosaic of agricultural paddocks and timber woodlots. Let the attractors be the geographic centres of each woodlot/paddock. Suppose we define a membership function that for each point is a vector pointing to the corresponding attractor. Assume for the sake of simplicity that we have only defined one geographic field over the territory, the tree cover fraction (TCF), in the same way as in previous examples. Now imagine we travel along a transect crossing several of these geographic objects. As we approach the boundary between a paddock and a woodlot, the membership function becomes unstable, eventually it disappears at the very boundary and soon after crossing it, it points towards another attractor. On the other hand, if we consider that a discontinuity in the TCF field occurs whenever its value changes abruptly (say from 0 to 100% in only one disk-diameter distance), it can be seen that the discontinuities in the field occur at the boundaries between paddocks and woodlots, coinciding with the singular points of the membership function. Therefore K is the set of singular points of W.

We shall call K the *morphology* of the phenomenon we wanted to foreground from the territory, 'forestness' in the example. K is the system of qualitative discontinuities formed by phase transitions in the system of attractors. The morphology K is the set of boundaries that makes forestness salient as a phenomenon. Furthermore, it constitutes what Marr (1982b) termed the ***primal sketch*** of an image in the context of computer vision, that is, the first step in transforming a numerical image representation (a discretised geographic field) into a symbolic shape-oriented representation (a thematic vector layer). The morphology's singular points are the zero-crossings (i.e. breaks in image intensity) in Marr's terminology. In chapter 3, a general method to obtain this primal sketch will be presented.

By now it should be clear that **the two views of geographic reality, the continuum (field) model and the discrete (object) model, are by no means mutually exclusive neither contradictory**. Rather, geographic objects are higher-level entities that are derived, at least implicitly, from fields. Since qualitative differences that separate geographic objects from each other are supervenient upon quantitative ones, the qualitative space made of individual entities that conforms our entity view of geographic reality is actually a partition of the quantitative space made of geographic fields. The field model is a more exact representation of the state of a territory, whereas the object model is more easily graspable.

In this section we have seen two explicit ways of producing such partition. The first one consists in defining precisely what a particular class is in terms of some quantitative thresholds specified over some relevant fields. The intersection of the resulting isolines in all the considered fields produces a partition of geographic space in which each region conforms to a class definition. The second way is to locate singular points (boundaries) in geographic space according to discontinuities in the set of relevant fields. Each region enclosed by the resulting morphology (the closed network of boundaries) is an individual object whose nature (class) has to be determined in a later stage. The first method considers class-concepts as *mass nouns* referring to homogeneous materials. The second one takes classes as *count nouns* referring to types of geographic objects. Therefore the latter is the method of choice for object-oriented analysis (further discussion of this topic in 2.7).

2.3. Model background: Frank's five-tier ontology

The model of geographic reality that will be expounded in the next sections is based upon the multi-tiered ontology for spatio-temporal databases proposed by the Austrian geographer Andrew Frank, first sketched in (Frank 2001f) and further developed in (Frank 2002). An *ontology*, as regarded in the information systems literature, is a set of formal definitions and relations of a collection of entities, i.e. an explicit account of a conceptualisation of a structured part of reality. Frank argues that GIS, in order to convey a useful description of the world, must be constructed, at least implicitly, on the basis of some ontology. This was not clear in the beginnings of GIS and as a result many practical problems arouse due to inappropriate ontological assumptions. In (Frank 2002), he investigates the minimal set of ontological commitments that underpin the relation between geographic reality and GIS, and sets forth the following:

1) <u>Existence of a single reality</u>: you and me take each other for granted, and live and interact with the same objects, following the same physical laws.
2) <u>Values for properties can be observed</u>: our knowledge of the world is obtained from measurements of observable properties, including inferences of properties not directly measurable.

3) Assume space and time: any observable property is necessarily linked to a spatio-temporal extension, therefore space and time have to be modelled in advance, as a system of reference in which measurements from observed properties are placed.
4) Observations are necessarily limited: as pointed out by Heisenberg's principle, measurements come necessarily with limited precision in the observed value as well as in the point in time and space they are related to.
5) Cognitive processes determine objects: for economy of cognition, sets of similar observations are grouped to form objects, which depend on the classification used for their formation.
6) Names of objects: objects endure in time and have an identity that is usually represented with identifiers (names) in the data collection, as to make them unique and easy to find without constant tracking.
7) Social constructed reality: the social system constructs virtual relations between objects based on general agreements and convention. These relations are only 'real' in the context of the social system that created them.
8) Knowledge of an agent is changing in time: agents (individuals and institutions) use their knowledge to decide between alternative courses of action, but this knowledge is incomplete, partial, and normally referred to a past time. Therefore the decisions of agents have to be judged with respect to what they could have known at the time they made such decision, as it is actually done in environmental and economical audits.

Based on these commitments Frank (2001f) proposed the construction of an ontology consisting of five coordinate tiers that can integrate different ontological approaches in an unified system:

Tier-0: autonomous physical reality. This tier assumes the existence of a unique physical reality independent of human cognition and that can be modelled as a four dimensional continuous field of attribute values $f(x,y,z,t)=a$, where each attribute has a unique value at each time/position.

Tier-1: observation of physical reality. The observation of physical reality with synoptic or local instruments enables the construction of the next tier, based on snapshot measurements of some selected attributes that are made at some locations and times with limited resolution and precision.

Tier-2: objects with properties. The former measurements are used as the raw material upon which objects are carved out. Objects are defined as spatio-temporal regions where the set of relevant properties are uniform. In order for objects to have uniform properties, the latter must be classified and small variations in the measurements, eliminated or obviated. Given a classification to determine what is a uniform object, one can delineate all objects within an area. The resulting objects form a partition, i.e. they jointly exhaust the space and are mutually disjoint. Objects are formed such that many important properties in them remain invariant through time, so that they can be allocated a stable identifier.

Tier-3: Socially constructed reality. Social reality includes all the objects and relations that are created by social interactions. The primitive of this tier is the formula 'x serves or counts as y in the context of z', which fills the former tier with a set of functional relations and surrogate properties that may eventually create new objects (e.g. a golf course) from the already identified ones (e.g. the lakes, grassland and woods that compound the former).

Tier-4: Cognitive agents. Cognitive agents (individuals and institutions) are capable of logical deduction. From the knowledge accumulated other facts are deduced and used to guide the actions of the agent. This tier encompasses the set of rules through which the database is queried. It should be constructed in such a way that the conclusions drawn from querying the database are the same as the ones that would be obtained by inspecting the world directly. One important problem of this tier is that time may affect the validity of the deductions, since they are based upon past snapshot information, so that, as the database is updated, past deductions have to be revised (this is known as the 'belief maintenance problem'). Another problem is to determine the reliability of each source of information and how this affects the reliability of deductions made combining different sources.

2.4. Model motivation and overview

The model of geographic reality expounded below is an attempt to **i)** justify the validity of the use of remote sensing images to produce spatial information on landcover, and **ii)** formalize the conceptual foundations previously stated into a model. The model provides a framework that states explicitly how the objects created by the analysis relate to the underlying real world. It constitutes the basis of a classification method that is oriented to the construction of geographic objects from its very initial steps. The main problem addressed is that the properties measured by remote sensors, as we thoroughly saw in the previous chapter, relate only indirectly to the properties used to classify landcover. The discourse will specifically refer to vegetation, since the map used as an example is the Forest Map of Spain (MFE, which is briefly described in appendix 4) although it may be applicable to general landcover or any other kind of area-class maps.

As Bennet (2001) expresses it, 'Although indirect methods may be effective for many purposes, they do not elucidate how to partition vegetation-types in terms of properties of the vegetation itself and hence, from an ontological point of view, they are suspect because they define something in terms of factors that are only contingently related to the phenomenon in question'. For this reason the model will be presented in two versions, one related directly with the biophysical attributes used to classify vegetation, and the other more accord to the pragmatic way in which landcover maps are produced. The main difference between both versions is their economic constraints, which are inexistent in the first one. The assumptions that allow the shift from one version to the other will be stated and investigated.

Both versions, hereafter called respectively the *idealistic model*, or ***I-model***, and the *realistic* or ***R-model***, consist of three tiers similar to the first three tiers of Frank's ontology. The construction of the two remaining tiers, tier-3, that models the uses, functions and other agreed properties that a society grants to its territory, and tier-4, that models the way agents use and exchange functional information from the spatial database, are beyond the scope of this thesis and therefore will be set aside from the discussion. Hence, the tiers that will be studied here are tier-0, or ***commonsense tier***, tier-1, or ***field tier***, and tier-2, or ***object tier***.

The first (commonsense) tier consists of the commonsensical reality that was taken for granted in 2.2.2. It differs from Frank's tier-0 in that the elementary gravitational, electromagnetic and nuclear fields enabling the cohesion, unity, permanence and other properties of the objects to which Horton's (1982) primary theory (2.2.2) refers, are considered of no interest for the purposes of geographic investigation. That is to say that the atoms of the commonsense tier are the mesoscopic material objects that we perceive and interact with in our daily experience. Hence here, in contrast to Frank (2002), who grants physical properties an ontological precedence over objects, the properties that are relevant for the study of landcover are taken as supervenient on the mesoscopic sessile objects that populate the Earth surface, like trees, herbs, boulders and buildings, which are the constituents of the geographic objects of interest. Under this view, the fact that e.g. a prairie is green is not explained in terms that there is a high concentration of molecules that absorb blue and red light and reflect green light, but simply that there is a high density of herbs, which are green. Note however that the ontological precedence is again reversed when we talk about geographic objects, which are supervenient on a set of properties defined by the corresponding geographic fields.

The next (field) tier consists of a set of geographic fields, defined in 2.2.20 as continuous regionalised variables, upon which the following (object) tier will be constructed. Prior to this, geographic location is modelled as a 2D space in which the abscissa and the ordinate are respectively the latitude and longitude of geographic points according to some cartographic planar projection as e.g. UTM. The reason for this choice instead of the more natural 3D space is that this system of reference is more easily handed by humans accustomed to paper maps and by current GIS. In the I-model, these fields correspond to the biophysical attributes used to discriminate between vegetation types, whereas in the R-model the fields are the digital images used in the analysis. In both cases, the fields are derived from the observation and measurement of the commonsense-tier.

The last (object) tier considered, consist of classified geographic objects, i.e. the map itself. In the I-model, they can be constructed in two ways: point-wise field classification or field segmentation. The first method yield objects that on the one hand, are delimited by isolines that may not coincide with genuine discontinuities, and on the other, may include important boundaries in their interior. In contrast, the objects derived from the second method fit better vegetation variations. In both cases, internal heterogeneity is increased when a MMU constraint is imposed. In the R-model, only the second method is used, and the resulting boundaries coincide with the ones of the I-model (when obtained with the same method) only when the luminance variations are produced by changes in some feature of the vegetation that is relevant for the classification scheme. The final configuration of the object tier in the R-model is constructed in successive steps, each one introducing a higher level of uncertainty. Since it will be assumed that the only way to keep at bay error propagation is to allow for direct human intervention in the process, the automated part will be restricted to the lowest meaningful level possible, that is, the lowest level where all the regions in the partition are potentially meaningful as individual entities.

2.5. The Idealistic (I-) model

The idealistic model of geographic reality assumes that there are infinite resources for the measurement of vegetation properties, i.e. that the MFE (appendix 4) project has an unlimited budget and personnel. In such scenario, we could measure in the same day all the relevant attributes of vegetation in each point of the 115x70 km^2 territory encompassed by each MFE sheet.

2.5.1. Field tier

Suppose we hire all the ecologists of the world, summing up a crew of one million people. We choose a clear late spring day, and early that morning we distribute them systematically in that rectangular territory along UTM easting lines spaced 100m, and at each ten metres of those lines we place a staff. We give each staff a differential GPS with <10cm positional accuracy, a metric tape, a hypsometer (to measure tree height), a data logger, and other survey material. We instruct them to establish ten circular plots of 12.62 m radius (500 m^2), centred at each ten metres of a UTM northing line in the space between the initial easting lines. In this way we would obtain a square ground sampling interval of 10 m which, given the size of the measurement disks and the MFE cartographic scale, can be considered as continuous over the whole territory.

Each staff must measure and record the following items within each plot (measurement disk):

1) Exhaustive floristic inventory, consisting in marking the name of the species present in the plot, from a list of some 5000 species from *flora iberica*[1].
2) Tree cover fraction (TCF, % of the surface of the disk covered by plant individuals >7m).
3) Species break down of TCF (distribution by species of TCF, up to four species, listed in descending order. Species with less than 5% CF are not included).

[1] A complete list of vascular plants living in the Iberian Peninsula and the Balearic Islands, either native or naturalised, compiled by the Royal Botanic Garden of Madrid (http://www.rjb.csic.es/floraiberica).

4) Shrub cover fraction (SCF, % covered by individuals >3m and <7m).
5) Species break down of SCF (distribution by species of SCF, up to four species. Species with less than 5% CF are not included).

i) ... same (CF and distribution) for each of the remaining height strata defined;

14) Type of ground not covered by vegetation (rock, sand, concrete, etc).

Note that the overall vegetation cover fraction is the sum of the CFs of the strata, since parts covered by a higher stratum are not considered in the calculation of a given lower stratum. Once the collected data are ingested into a GIS, we could obtain thousands of continuous fields, each one describing the abundance (%CF) of individuals of a given species in a given height stratum. The number of fields per stratum is in general inversely proportional to the height of the stratum, since there are less tree species than shrub species, less shrub species than bush species, and so on. The fact these fields become zero at some location does not warrant the absence of the species, it just means that CF of individuals of that species having that height interval is less than 5%. Therefore, if we want to know the actual geographic distribution of a given species, we would have to resort to the dichotomous (having two values: 1, present, 0, absent) fields derived from the floristic inventory.

2.5.2. Object tier

Once the set of relevant fields has been constructed, there are two alternative ways to foreground the objects of interest: either vegetation is classified at each point and subsequently the objects are formed connecting points equally labelled, or the contours of objects is first delimited by locating discontinuities in the fields and then the content of each object bounded by such discontinuities is determined.

2.5.2.1. Point-wise field classification

The first step in the classification would be an algorithm allocating a label to each point. In MFE, labels consist primarily of up to four initials of the main species found in the polygon, according to the rules explained in appendix 4. The algorithm could be as follow:

For each measured point in the territory:

1) Put the initials of the species found in item 3 (TCF distribution) in the point's label, following the same order.
2) If there are empty positions left in the label, and TCF < 35%, then put the initials of the species found in item 5 (SCF distribution) in the point's label, beginning by the first empty position and keeping the same order.
3) If there are empty positions left, and (TCF+SCF) < 35%, then put the initials of the species found in item 7 (BushCF distribution) in the point's label, beginning by the first empty position and keeping the same order.
4) Repeat this process for the remaining strata, whenever there are empty positions left and the sum of already processed CFs is <35%, otherwise end.
5) If after 4) the label is empty, put the type of bare ground as the label.

Note that the four available positions of the label need not to be all filled. A label may consist only of one species, if e.g. it is arboreal and it covers more than 35% of the disk. Once each point has a label, the next step is to form homogeneous objects or regions, each one consisting of a set of connected points that share exactly the same label. Within this context, two points are connected if there is path between them that does not cross points differently labelled. The resulting regions are jointly exhaustive and mutually disjoint, i.e. they form a partition of the territory.

These classified objects have **four problems**. **First**, they are delimited by isolines (in the MFE, by the 35% CF threshold) that may not correspond to genuine boundaries. Imagine e.g. a forest whose density decreases gradually from its core, from 100% TCF to 0%. Then a wide strip of the forest will be left out of the 35% TCF isoline. **Second**, they may include in their interior vegetation boundaries that may be significant for the users. Imagine e.g. two adjacent Scots pine stands, one uniformly dense (say 100% TCF) and the other uniformly sparse (say 40% TCF). A forester would probably consider them as distinct units, yet they form a single object in the partition. **Third**, their configuration may convey a sense of heterogeneity in places that are considered uniform by users. Imagine e.g. a dense mixed pine/birch forest where TCF for each one of them is roughly 50% at every point of the forest. Points with a slight preponderance of pines will have a different label than points where birch prevails. These tiny differences split the forest into smaller objects that may be negligible for users, who may consider the forest as a unitary uniform whole for all practical purposes. And **fourth**, they may be too small to be significant for users.

While the first two difficulties are unsolvable using this labelling system, the third and fourth can be tackled with relative ease. Once a MMU size is selected, all the regions smaller than MMU will be merged to the most similar neighbour. Similarity has to be defined in both floristic and physiognomic terms, using e.g. one of the many existing similarity indices (see e.g. (Legendre & Legendre 1998)). By setting a sufficiently high threshold, we could also merge adjacent regions whose similarity is above that threshold, solving in this way the third problem. The label of the merger would be the most frequently found in points within the new object. The fourth problem is tackled by the introduction of a MMU, although in zones where there is a predominance of small regions, they would be better represented within a larger heterogeneous 'mosaic' object. In this case, the choice of regions to be included in the mosaic, as well as its final label, may not be trivial and require some kind of human intervention.

Finally, the other MFE attributes out of the label have to be defined for each region. The height interval can be set as the one most frequently found in points belonging to the region. The spatial distribution of each relevant species is a more complex issue, requiring the use of some geostatistical technique (as e.g. variograms, see 2.2.23) in order to assess it. The same would apply for the spatial distribution of facets within mosaics.

2.5.2.2. Field segmentation

Within this context, segmenting a field implies i) detecting the stable attractors (points of minimum variation) of the field, and ii) contouring the area of influence of each stable attractor. The boundaries of the areas of influence define the morphology *sensu* Thom (2.2.29) of the field, that is, a complete partition of it. The aim of this method is to detect discontinuities in the set of fields that are relevant for classification, i.e. discontinuities implying changes in the physiognomy and/or floristic composition of vegetation. A way of doing this is to proceed in a sequential manner, as follows:

1) segment the individual species TCF fields.

2) overlay the resulting partitions as to obtain a single partition. After this, each region has a uniform value in each attribute that is different (in at least one attribute) than the one of its neighbours.
3) mark all segments where the overall TCF > 35%. Those segments have already a final MFE label and therefore can be preserved for the final partition.
4) segment the individual SCF fields.
5) overlay the resulting partitions as to obtain a single partition.
6) erase all the boundaries overlapping with the areas enclosed by marked segments of output partition from 2). By doing this, segments already having a final MFE label are preserved without further division.
7) overlay the output partition of 6) with the one of 2).
8) mark all segments where the overall (TCF+SCF) > 35%.
9) repeat this process for the next stratum.
10) stop when all the territory is marked or when the last stratum has already been processed.

The final partition consists of segments, or patches, of homogeneous physiognomic structure and floristic composition that differ from their neighbours in at least one relevant attribute. By definition, they are bounded by genuine discontinuities, and their interior is free of significant boundaries, an important property that the regions from the former method lack. However, this method may show a problem that field thresholding lacks, which occurs when there are no discontinuity but gradation between two regions that are semantically different, like e.g. a forest of species *x* that changes gradually to a forest of species *y* along an altitudinal gradient. In this case, the boundary between both forests is not a thin strip but a wide area. Nevertheless, it can be expected that **the transitional area is not free of granularity**, so that there will be places within it where the rate of either floristic or physiognomic change is high. Such places will be identified as boundaries, and the result is that the transition zone is likely to be partitioned into a series of smaller patches where none of both species can be said to be predominant, and therefore may be classified as 'mixed x-y forest'.

In order to classify the segments, the mean value in each attribute of the points within the segment is computed. Afterwards, the same labelling algorithm of the previous method is applied, using the segment mean values instead of individual point values. As in the former case, some segments may be too small to be significant, so that an MMU has to be applied together with a similarity measure. Merging rules and aggregation into mosaic objects can be performed as in the previous method. The same holds for the remaining MFE attributes. Note that, even applying the same rules and MMU, the final partitions of both methods will not coincide. In general, the second one will approximate better landscape variations, since every boundary in it correspond to a discontinuity, in contrast to the point-wise field classification method, that may include a considerable proportion of arbitrary boundaries.

2.5.3. I-model summary

The I-model assumes a scenario where there are unlimited resources for the measurement of properties of geographic phenomena like vegetation. These properties, which are supervenient on the sessile objects (plants) that compound the first tier of the model (commonsense reality), are quantified by a set of precisely defined attributes. The latter are measured directly from the ground at every position of the territory under study, which is modelled as a 2D space. The measurements yield a set of continuous exact fields (called *I-fields* in the next subsections) constituting the second tier of the model. In the example, they were constructed using a common disk size, although in more general landcover maps, different disk sizes may be required for fields not related to vegetation. In the last tier, the geographic objects of interest are constructed using the former fields. There are two alternative ways to foreground classified objects: either the geographic phenomenon (vegetation in the example) is classified at each point and subsequently the objects are formed connecting points equally labelled, or the fields are partitioned by locating discontinuities and then the content of each region is determined and conformed to a classified object. In general, the second method approximates better landscape variations. Both methods lead to a complete partition of the 2D space representing the territory, in which there will be too small objects that have to be merged into larger objects. This is done by selecting i) a MMU size below which objects have to be aggregated, and ii) a similarity measure to decide to which neighbour each small object will be merged.

In the I-model, the set of relevant attributes is perfectly known at any point of the territory. Assuming there has been no error in the measurement neither in the recording, this knowledge is certain, since it has been obtained from direct mensuration. If we retain only the idealistic object tier, uncertainty is inevitably introduced, since point-wise information is lost. However, we still have certainty about the actual range of values that the attributes take within each object. As the MMU is introduced, the uncertainty is increased provided only the final partition is stored. In this case, we cannot know whether an object has some parts that are out of the range given by the database. However, we can assume the extent of such 'anomalies' can be neglected, otherwise the object would have been categorised as 'mosaic', and additional information on the different parts compounding it would be available. In mosaics, we know the areal percentage occupied by each type of facet and their general pattern of distribution within the mosaic, but we cannot know where each particular facet is located.

2.6. The realistic (R-) model

In the realistic version of the multi-tiered model of geographic reality, the available resources for the compilation of the map are limited. In this case it is unfeasible, due to technical and/or economical reasons, to measure extensively many of the relevant properties intended for classification. The solution generally adopted is to operate on a set of surrogate fields given by remote sensing imagery, complemented with a limited field survey.

2.6.1. Field tier

Such fields, which are overwhelmingly cheaper and easier to acquire than the ones of the I-model, are derived from measurements drawn not from the ground but from remote sensors. The latter convert the incoming radiation into evenly spaced measurements that make up a digital image. Thus the attributes measured (optic radiance, radar backscatter, lidar waveform, etc) relate only indirectly to the properties used to classify landcover (see 1.9 for a thorough discussion). This relation is obscured by external factors such as atmospheric and illumination condition, soil moisture, etc. In order to simplify the discussion, we are going to assume in a first approximation that those factors are negligible or can be adequately corrected. This premise adds to the already stated one that the positional accuracy of the data is good enough, i.e. that the plot of terrain to which each pixel refers can be precisely located.

In the R-model, the **primary assumption** enabling landcover mapping with EO data is that **the overall spatial variation of the RS-derived fields coincides to a great deal with the one of the relevant I-fields.** The validity of this assertion, which hereafter will be called the *coincidence hypothesis*, is discussed below. The adjective 'overall' means that the set of images used in the analysis will be considered as a single 'colour' or multiband image. It assumes that the images are highly redundant, i.e. they show the same structure but with different emphasis. In other words, luminance variations occur at the same locations in all the images, it is only their magnitude what changes from one image (band) to another. What the coincidence hypothesis states is that this variation is accompanied by a change in at least one of the relevant idealistic fields representing the properties used for classification.

More precisely, suppose we define a quantitative dissimilarity measure between pairs of geographic points according to the value of the I-fields at those points. If we compute this measure for each pair of adjacent points (where adjacent means that the distance between the points is the disk diameter, which in turn should have the same size than the sensor GIFOV), we could calculate, using two orthogonal directions, the magnitude of the maximum variation (gradient) at each point. The result, once discretised, could be displayed as a digital grey-level image representing gradient magnitude of the dissimilarity measure. Now assume we do the same on the RS image, i.e. define a radiometric distance so that another gradient magnitude image is derived. What is hypothesized is that both images will show a conspicuous resemblance that would result in a high correlation coefficient. But before discussing its validity, it should be explained why RS ortho-images can be considered as geographic fields.

A RS digital image consists of a set of regularly spaced discrete measurements. Once ortho-rectified into some planar cartographic system of reference (so that each pixel corresponds to a square terrain plot of known coordinates), it can be easily transformed into a continuous geographic field. The analogy consists of measurement disks placed at the centre of each pixel, with a diameter equal to the width of the GIFOV (see 1.4). All detectable photons coming from the disk are integrated modulo PSF into a single discrete value, which is the field value at the central point of the pixel. If we want to obtain the value of the field at any other non-central position, the actual DN format of the image, usually *byte* (8bits), has to be first converted into *floating point* format (32 bits digital numbers with up to 7 decimal positions). Then, the value of the field at a given coordinates is obtained by interpolating the value of the four spatially closest central points. Since the ground sampling interval (GSI, or pixel size) usually matches the GIFOV, there is no point of the territory left unobserved, i.e. all the points contribute to the value of at least one pixel. In this situation, it can be safely assumed that the value estimated by interpolation at a given geographic point is close to the actual value that would have been obtained if a measurement disk were placed at that point. Hence, it can be assumed that the digital image is a continuous regionalised variable measuring a precisely defined attribute like e.g. near-infrared reflectance, that is, a geographic field as defined in 2.2.20.

Once we have granted ortho-rectified RS images the status of geographic fields, suppose we survey the RS-derived partition extensively on the ground. If the coincidence hypothesis were invalid for most parts of the territory, we would learn that any coincidence between the boundaries of the partition and the spatial distribution of landcover within the territory is fortuitous. Since evidence from mapping accuracy assessment does not point towards this situation, it can be affirmed that there is some empirical ground to believe that the coincidence hypothesis holds in many situations. In fact it is the isomorphism between the images and the features of the territory observable at that scale what preserves the validity of the hypothesis. When the property measured by an idealistic field is observable from the air, we are likely to find an image where spatial change keep a high correlation with the one of that field, as e.g. the vegetation cover fraction and the thermal infrared band. On the other extreme, there will always be some I-fields that cannot be observed from the air, like e.g. the distribution of a particular understory species. In this situation, the coincidence hypothesis would hold only if those fields are not functional, i.e. they add descriptive information but are not taken into account in the classification scheme, as it actually occurs in MFE.

Apart from the already obviated atmospheric and illumination effects, there are two general cases leading to a local violation of the coincidence hypothesis. The first one occurs when a change in some relevant idealistic fields (e.g. canopy species composition) cannot be detected in any of the available images (e.g. the boundary between a *Pinus pinea* and a *Pinus pinaster* stands). The second one is the opposite situation, when a change detected in most of the images corresponds to a change in a ground feature (e.g. soil) that is not considered by the classification scheme. While the latter case produces spurious boundaries that can be identified and erased with relative ease, the former omits semantically important boundaries that are difficult to delineate *a posteriori*. Two solutions seems acceptable in this case: either the taxonomic resolution of the classification is lowered (so e.g. both contiguous pine stands form a semantically uniform object), or the resolution is maintained as it was but it is explicitly accepted that, assuming the field survey will detect the presence of two pine species in the object of the example, there may be a special case of mosaics (2.2.27) where the percentage occupied by each type of facet is unknown.

Another issue affecting the coincidence hypothesis, related to the Change of Support Problem (2.2.26), is the ratio between the GIFOV and the size of the measurement disks of the idealistic fields. The GIFOV is the regularising element in RS images. The greater the GIFOV, the less detailed account of the territory can be obtained from the image. In this sense, the GIFOV is a generalization mechanism similar to cartographic scale. Regarding the measurement disks, we saw in 2.2.18 that their size is chosen so that the resulting measurements are homogeneous over a zone that is considered homogeneous as regards the property to which the idealistic field refers. If the GIFOV is several times greater than the selected disk, many of the differences encountered in the I-field will be smoothed out in the image, so that the correlation between them decreases. Conversely, when the GIFOV is several times smaller than the disks, there will be a lot of discontinuities in the image that have no counterpart in the I-field, so that the correlation is again lowered. Therefore, given a particular type of image, the coincidence hypothesis will be more plausible as the GIFOV approaches the size of the selected disk.

Regarding the latter, it should be noted that not all the I-fields may have been constructed using the same disk size, as explained in 2.2.22. For example, a field measuring building density would need a disk larger than the one required for constructing the TCF field. The consequence is, as pointed out by (Marceau, Howarth, & Gratton 1994), that 'there is no unique spatial resolution appropriate for the detection and discrimination of all geographical entities composing a complex natural scene'.

Finally, I would like to mention that an important question regarding RS-fields is left unanswered in this thesis: what is the best combination, in type and number, of images that can be used to analyse landcover, from a given data set. This is a classic problem of pattern recognition, called *feature selection* (where *feature* stands for attribute), consisting of two inter-related parts: feature extraction (the transformation and/or combination of the original images into new ones) and feature reduction (the reduction of the dimensionality of the data set by selecting the smallest subset of features providing an acceptable discriminative power). Feature selection is generally considered a process of mapping the original measurements into more effective features. Unfortunately, in many applications, the important features are non-linear functions of original measurements. Since there is no general theory to generate mapping functions and to find the optimum one, feature selection becomes very much problem oriented (Fukunaga 1972). In any case, the two main approaches used (Mausel, Kramber, & Lee 1990) are class separability analysis (with e.g. the Jeffries-Matusita or any other *ad hoc* distance) and evaluation of eigenvector and eigenvalue data derived from class statistics (e.g. canonical discriminant analysis). In order to avoid this issue, and given the snapshot nature of landcover maps, the data set will considered as made of only one date and type of images, typically consisting of either a panchromatic very high resolution (VHR, <5 m) image or a high resolution (HR, 10-30 m) multispectral image.

2.6.2. Object tier

In the R-model, likewise the I-model, there are two alternative ways to foreground the objects of interest: either the RS-derived fields are classified at each point and subsequently the objects are formed connecting points equally labelled, or the objects are delimited through some segmentation technique and afterwards the content of each object is determined.

In 1.10 we saw how typical pixel-based classification is performed. It consists in delineating the regions of the multidimensional data space associated with each class of interest with the aid of discriminant functions. This approach is not directly applicable in the case of the MFE, since there are no formal classes defined (no two polygons are exactly alike in all the attributes considered). Even if we only take the floristic composition appearing in the label, the number of possible combinations greatly exceeds the possibilities of any classifier.

Considering also that the aim of this chapter is to set forth the foundations of object-oriented classification, the pixel-wise approach will not be addressed here. However, it is worth noting that a classification scheme could be constructed in the case of MFE if a hierarchical clustering of the set of all MFE polygons is carried out using a combined floristic and physiognomic similarity measure, as proposed in 2.5.2.1. In order to establish the basic level of the classification, the resulting dendrogram would be cut at a height where the number of clusters is relatively small. These clusters would constitute the classes of the scheme after having been named. Another simpler approach would be to assign each polygon to a class of an already established scheme as e.g. the NVCS (TNC, 1994).

2.6.2.1. Image segmentation

Segmentation, as understood within this thesis, is the process of deriving, based upon the spatial structure of the image, a partition of it into a set of jointly exhaustive, mutually disjoint regions, or *segments*, having the following properties:

> I) the partition is a representation of the structure of the image, where most of the spatial differences of the latter are obviated, but the ones retained are accurately represented, so that the boundaries separating the segments correspond to genuine discontinuities in the image;
> II) all the segments exceed the minimum size imposed by users as to be potentially meaningful for them as individual entities; and
> III) all the segments are relatively homogeneous, so that their degree of homogeneity is higher than the one that would have the union of any given segment with any of its neighbours.

Segmentation is the first stage of object-oriented image analysis. There are hundreds of segmentation algorithms that have been developed in computer vision and medical radiology (see e.g. (Cufi et al. 2002) for a review). They determine the shape, size and location of a series of non-overlapping regions that are later identified as objects (or parts of objects) constituting instances of some classes. The object-oriented approach conceives classes as *count nouns* (e.g. woodlot), therefore the areal units on which it operates must exceed the minimum size required for objects to be recognisable, i.e. as to identify them as class instances. In contrast, pixel-based classification assigns labels to individual pixels and later forms objects by connecting pixels equally labelled. In doing so, it conceives classes as *mass nouns* (e.g. timberland) referring to homogeneous materials, so that the area enclosed by a single pixel can be an instance of a class.

2.6.2.1.1. The first partition as the primal sketch of the structure of the image

Using graph theory[1], an (either single or multi band) image can be conceived as a snapshot from a planar dynamic network consisting of triangular meshes made up of nodes (pixels) connected through links via which the nodes interact. The interaction consists of quantitative luminance exchanges between the nodes. The intensity of the interaction is regulated by proximity in the data space, decreasing rapidly with radiometric distance, and it is formalised through a *weight* allocated to each link. A link may be *active*, if there is some noticeable interaction through it, or inactive if its weight is nearly zero. Each node is connected by links with its eight immediate neighbours, and has an associated vector indicating the overall magnitude and direction of the interaction. Two nodes are said to *interact* if there is path of active links connecting them. A node is called a *local attractor* if the interaction vector of each of its eight neighbours is pointing towards it. A node is called *regular* if it interacts with all its neighbours. The area of influence or *basin of attraction* of a local attractor is the set of regular nodes interacting with it. A node having some link(s) inactive is a *singular* node. The set K of singular nodes forms the *morphology* sensu Thom (1975) (the boundaries, or salient parts, of the structure of the image separating the functional units, or basins) of the network.

The structure of the network is given by the signature (luminance vector) associated to each node, which in turn determine link weight. The original image corresponds to the structure of the network at time t=0. Variable t is not physical time but a cumulative discrete variable counting interaction cycles. In each cycle, the signature of the node may change as a consequence of the interaction, therefore the structure of the network is dynamical. Two nodes that interact tend to approach each other as well as their common attractor in the data space. If we allow the network to evolve during a sufficiently long period of time, some nodes will interact more strongly between them and some will stop interaction with some others, inducing a coherent behaviour of the nodes within each basin. Eventually some inactive links may become active, opening paths between nearby attractors. As a result the weakest attractors and their basins will be captured by stronger ones. This process leads to a decrease of the number of attractors and to an increase of the homogeneity (proximity in the data space) within each basin of attraction, that is, a simplification of the structure of the network.

[1] The mathematical study of networks and topological maps. An introduction to graph theory can be found in (West 2000).

Such simplification can be viewed as an evolution towards a piecewise constant image in which the pixels within each basin have roughly the same value.

After relatively few cycles, the network reaches a steady state far from equilibrium (where equilibrium would suppose a uniform distribution of luminance across the network, i.e. a flat image). In the steady state, the interactions are balanced and the network structure remain stable, i.e. change at every node is negligible. The remaining attractors within the network are called *stable attractors*, and their respective basins of attractions will be called *blobs*, or *primal segments* (defining Marr's (1982b) *primal sketch*, see 2.2.29), that are considered the primitives of image structure. Moreover, blobs are not only structural basic units, but also functional, under the analogy of the dynamic network. The term *blob* is a perceptual concept used in image analysis to refer to a homogeneous small region, darker, brighter or of a different hue than its surroundings. By adopting such name it is assumed that the basins of stable attractors are perceived as blobs in the image. It also helps to distinguish them from other type of regions, to be created later, consisting of aggregations of blobs, which will be termed generically *segments*.

Note that the partition defined by the morphology is a rough zonation in the sense of 2.2.14, since singular pixels do not belong to any of the blobs they separate. The analogy is that singular pixels are the cells of the reference grid where partial overlap takes place, and the basins of attraction are the full overlap regions of the blobs (see 2.2.13 for a description of overlap functions). To finish up, the link between the network and the method presented in chapter 3 is as follows. The first basic operation of the method is called ***steady_state***, which departing from the original image filters it out until convergence is reached. After this, the next operation is ***get_morphology***, which locates the singular nodes bounding the basins of the stable attractors. A third, trivial operation is ***label_blobs***, which assigns a unique numeric label to each blob.

2.6.2.1.2. The size constraint

So far we have a partition of the image fulfilling conditions I) and III), but not necessarily II). The question is how the blobs should be aggregated until they reach the minimum meaningful size. But before trying to answer this question, the need for condition II) should be justified. We depart from the isomorphism between the RS-derived fields (atmospheric and illumination corrected orthoimages) and the features of the territory visible in them. Such correspondence (which also preserves distance) is the base of the more specific *coincidence hypothesis* between the structure of these orthoimages and the one of the geographic fields representing the actual value of landcover attributes. At the object tier, **what the coincidence hypothesis implies is that a structural/functional unit in the image corresponds to a structural/functional unit in the landscape, i.e. that each blob corresponds to a patch**.

The problem is that, due to the hierarchic nature of the landscape, the concept of patch is scale-dependent (Wu & Loucks 1995). In a broad sense, a patch refers to a spatial unit differing from its surroundings in nature or appearance (Wiens 1976). Patches can be characterized by their size, shape, content, duration, structural complexity and boundary characteristics (Wu & Loucks 1995). Therefore, on different levels, a patch may be from the area covered by an isolated tree to an island continent. Nevertheless, we have already defined what a landcover patch is within this thesis: a contiguous area of similar dominant species and physiognomy (height and cover) occurring in an area of similar physiography (aspect and slope). The upper bound of the possible extension of instances of this precisification is difficult to establish (it could reach hundreds of sq. kilometres, think e.g. of a savannah plain). However we can safely place the lower bound somewhere between a few thousands sq. meters and a few hectares, depending on our objectives and the complexity of the territory under study.

This lower bound is the minimum mapping unit (MMU). Patches smaller than MMU will be assumed in most cases to belong to a lower level, of little interest for our analysis, within the patch hierarchy, like e.g. the gap created in a forest by a fallen old tree. Less often, they could be 'proper' patches, but in this case it is assumed that their size makes them insignificant, i.e. that the fact that their presence is neglected does not change at all the picture. In other words, users informed of the presence of these small gaps would make the same decisions than users who do not know of them. The underlying premise of this reasoning, hereafter called the ***size hypothesis*** is: **if blobs correspond to significant patches of this level or higher, their projection on the ground must exceed the MMU size**. If in a forest/non-forest map we set the MMU to 10 ha, we are saying implicitly that any isolated group of trees of less extension than this may be a coppice, a woodlot, or a stand, but it does not qualify for the label 'forest'. Or stated in another way, at this level of generalisation of the territory, we do not care about isolated coppices, woodlots or stands. The *size* and the *coincidence* hypotheses, together with the *correspondence* one that will be introduced later, constitute the proposed pillars underpinning the object-oriented approach to landcover mapping.

The size hypothesis is violated in two general situations. The first one is when the MMU size is chosen only attending to budgetary constraints and it exceeds the mean size of some important patches (recall the small woodlot example in 2.2.25). The second one is when there are objects whose individual extent is negligible (say tens of times smaller than MMU) but their nature and abundance make them significant when taken together. For example, a protected forest progressively invaded by scattered illegal family houses cannot be properly monitored with this approach. The first situation could be tackled by a sensitivity analysis on the effect of MMU size (e.g. plots of MMU size against timber stock) that eventually could suggest the use of a minimum size smaller than the representational MMU. The second one would require a specific analysis aimed at detecting point-wise (at the scale of observation) disturbances of special significance. This could be carried out by a pixel-based (with the pixel size close to the size of the houses to be detected) classification with only the class of interest (e.g. roofs) followed by a moving window counting the number of pixels positively classified.

2.6.2.1.3. The final partition as the baseline to object-oriented classification

Regarding the scale of observation, as we saw in 2.7.1, the correlation between spatial variation in orthoimages and in I-fields is maximal when the GIFOV matches the size of the measurement disk. So in order to optimise the fulfilment of the coincidence hypothesis, the former operations should be performed on imagery with a pixel size similar to the disk diameter. The problem is that different I-fields may have a different optimum disk size (it increases with the mean size and spacing of the objects –trees, buildings- to which they relate), and therefore we would better analyse the images at different resolutions. Another reason for the use of a multiscale approach is that the optimum scale of observation is dependent on not only attributes but classes, so that the boundaries of e.g. a sparse woodland are better detected in a coarse image than in one where the individual trees can be seen. In other words, '**boundary distinctness is scale dependent**' (Hay et al. 2001). Finally, a multiscale approach is necessary because patches within each level are generally sufficiently variable in size that the different levels in the hierarchy overlap in size (Woodcock & Harvard V.J 1992). Vg in a VHR image some blobs may consist of grassy patches, while some others will simply be tree crowns.

In short, the hierarchical nature of the landscape, together with the great variability in size and appearance of the patches within each level of the hierarchy, make untenable the hypothesis that, given an image with a fixed pixel size, each image blob correspond to a patch of the same hierarchic level. Since there is no single optimal scale of observation, the only possible solution is a multiscale approach that follows the nested hierarchical model (Woodcock & Harward 1992).

In order to construct such model, one possible alternative would be to, departing from a relatively fine resolution, coarsen the image, detect the blobs of the coarsened version, and link these blobs to the original ones, so that original blobs overlapping with the same 'coarser' blob are merged. This cycle (i.e. coarsening again the image and linking blobs across scales) would be repeated until all the mergers are larger than MMU. However, linking blobs across scales is not as straightforward as one might think, since the evolution of blobs through a stack of coarsened images does not only consist of merging events, but there can also be annihilations, creations and even splits (Lindeberg 1993). Therefore there may be situations where the assignment of previous blobs to new ones cannot be clearly established. These matching ambiguities between successive images, are known in computer vision as the correspondence problem (Cox 1993).

Instead, a simpler, faster method will be proposed in the next chapter, consisting in merging iteratively (only one merge per segment and iteration) blobs to their most similar neighbour until all the mergers are larger than the MMU, where the similarity is estimated using the values of the original pixels. The process can be seen as a multiresolution segmentation, assuming that each segment (recall that a segment is an aggregation of blobs), if represented by its mean, constitutes a resolution cell over which the signal is regularised (averaged). Since at each iteration the partition consists of fewer segments, the successive partitions can be considered a multiresolution representation (that is, a hierarchic zonation) of the original image, where each partition is an increasingly coarser (simpler) representation of the structure of the original image. In my experience, **it is more efficient to equate scale to the number and size of regions than to actually alter the pixel size (pyramid methods) or blur the image (scale-space methods)**. The latter have a deeper theoretical grounding (Lindeberg 1994), but region merging has also a sound basis, as explained below.

In order for output segments to fulfil condition III (internal homogeneity) of proper partitions in relation to the territory, another assumption, hereafter called the ***correspondence hypothesis***, has to be done. Such hypothesis states that **if two <u>adjacent</u> segments are radiometrically similar, they are semantically similar too**, i.e. their ground projections are likely to be parts of the same object. Note that this assumption is far less demanding than the one of the spectrometric approach regarding the existence of consistent, separable class signatures. It is restricted to neighbouring regions, and it relies (likewise the coincidence hypothesis) on the isomorphism between the image and the visible features of the territory. The spectrometric approach on the other hand assumes that spectral-semantic similarity is invariant to spatial distance, i.e. that the correspondence is unaffected by the First Law of Geography. Contrariwise, the hypothesis here assumes that the correspondence will decay with distance, and hence it is only applied locally.

Likewise the coincidence hypothesis, this one is violated when the type of image and/or the nature of objects from different classes make their signatures be located close to each other in the data space, showing the same (narrow) range of values in all the bands considered. In this case, blobs from these classes would be liable of being merged into a single segment whenever they appear adjacent in the territory. Again, the solution to this problem is to either reduce the level of detail of the classification scheme until the classes involved can be considered the same, or to accept that, assuming the field survey will detect the mixing, there may be a special case of mosaics (2.2.26) where the areal proportion occupied by each class is unknown.

During the segmentation process, not only the mean size of segments is enlarged, but also the size range is reduced, the lower extreme of this range being increasingly closer to the MMU size. Thus the representation given by the final partition would be equivalent to an image acquired by an imaginary sensor with an IFOV close to MMU that is able to adapt its shape to the structure of the imaged scene. Such sensor would deform and expand the shape of the IFOV until it finds a discontinuity, regularising the response of the surface bounded by such discontinuities. In some sense, the final partition is a representation of a particular level the *deep structure*[1] of the image given by a probe or window of the size of MMU, where all objects smaller than this size are traced over. Hereafter it will be named the ***baseline partition***, since it is the starting point of object-oriented classification in the R-model.

[1] The concept of *deep structure* was introduced by Witkin (1983) within scale-space theory to refer to the structure of an image at all levels of resolution simultaneously.

2.6.2.1.4. Granules

The segments forming this partition will be termed ***granules***. A granule is the minimal cell of a zonation that corresponds to a minimal cell in an associated taxonomy (recall the theory of granular partitions in 2.2.6). As such, it has no disjoint parts and no further parts, i.e. it is completely homogeneous (one meaning only) from the point of view of the classification scheme (taxonomy). It is assumed that the projection of the granule onto the ground correspond to a patch of the level defined in 1.4., or to a part of a patch of this level, where the rest of the patch is formed by nearby granules of the zonation. Granules have enough extension to be meaningful as individual entities (i.e. to constitute an object of interest) for users, whereas blobs may not. Depending on its neighbours, a granule may be categorised as an instance of some class (if e.g. it corresponds to an isolated woodlot), whereas a single pixel can never be an instance of an object class. In other words, every granule may become a mapping unit after classification. Therefore, granules are the *minimal semantic units* that can be established under the object-oriented approach given a predefined MMU size. The method to obtain granules from the initial partition, **build_granules**, is described in the next chapter.

The piece of terrain onto which a granule (roughly) projects may have some parts (gaps, see 2.2.25) that if larger, would have been considered as instances of another class in the taxonomy, but because of their reduced size, they are not recognised by the zonation, and thus they are traced over by the taxonomy too. In fact, the hardest arbitrary constrain imposed by this approach is the MMU size. Since every mapping project has to set, in order to tackle *sorites* vagueness (2.2.8) from gaps (2.2.25), a minimum size below which any ground feature is negligible no matter its nature, the arbitrariness is unavoidable. The consequence is that the Modifiable Areal Unit Problem (MAUP, 2.2.27) will appear. As it cannot be eluded, the best way to minimize its effects is to work at the lowest possible level of aggregation (Goodchild 1992), i.e. the lowest level where all the existing regions are potentially meaningful as individual entities.

This is one of the two reasons why segments are not allowed to 'grow' further once they reach the MMU size. The other is related to the fact that an error produced at an early stage of the modelling process propagates to the following ones, worsening the magnitude of the total error. Aggregating adjacent granules into mapping units (polygons) requires further information (coming from field surveys and previous models –maps) in addition to the one embedded in the images. Thus any merging decision beyond this point, if based solely on the images, is liable to produce an error that could have been avoided should the other information be used. Besides, **the correspondence between semantic and radiometric similarity is expected to decay with (geographic) distance**, so the larger the segments, the less likely the correspondence hypothesis.

It could be argued whether it would be more appropriate, from the point of view of error reduction, to use blobs instead of granules as the basic units. Blobs are classifiable from the spectrometric approach perspective (providing pixels are larger than the classification disk), but not all of them can be classified using an object-oriented approach. Rather, most of them will have a size inferior to MMU and therefore cannot constitute instances of object classes. In this sense, **the MMU size is equivalent under the object-oriented approach to the spatial resolution of the classification**. Notwithstanding it, the arrangement and properties of blobs within a granule may be used to describe the internal structure of the latter. In conclusion, if the *size hypothesis* is valid, blobs smaller than MMU would correspond to patches of little interest for users when considered as individual entities. However, in situations where the size hypothesis may be violated because of a large MMU imposed by economic reasons, it could be more appropriate to establish a minimum granule size somewhere between the classification disk of a spectrometric approach (with classes conceived as materials) and the MMU. This would enable, at the expense of a higher processing cost, the detection of some 'mosaic' polygons that otherwise would have been considered 'pure'.

Apart from the former reasons, there is a pragmatic point for using units larger than MMU. Currently there is no computer program able to emulate the capabilities of human interpreters, mainly due to the difficulty of modelling a highly complex process like vision. Consequently, it would not be unreasonable to expect that, in order to match the quality of photointerpreted maps, some kind of human interaction is required in the higher-level parts of a (semi) automated mapping procedure. Such interaction would be eased if the areal units presented to the user (as e.g. to confirm aggregation) have already a size that, on the one hand, makes them readily interpretable, and on the other, reduces the man-hours to an economically feasible period of time. In the end, the decision in favour of granules simply follows the parsimony principle commonly applied in science, that could be adopted as another law of naïve Geography: '**geographic objects are not to be multiplied beyond necessity**'[1].

To finish up the account on image segmentation in the R-model, the range of admissible pixel sizes to derive the baseline partition is discussed now. The finest resolution can be as high as desired. However, an ultra high resolution, although always useful for identification, may be of little interest for drawing patch boundaries of this level, so that below a certain threshold, it only increases the computational burden without improving the results. In addition, the fractal nature of patch boundaries will unduly increase the edge density (see 3.11). Then the minimum threshold could be set a little below the positional accuracy demanded for the final product (digital map), providing imagery of such resolution exists. On the other hand, the coarsest resolution should be finer than half the width of the MMU, otherwise patches of interest close to MMU would be liable to rest undetected.

2.6.2.2. Granule classification

Once we have the granules roughly located on the territory, the next task, which is beyond the scope of this thesis, is to locate them within the taxonomy. So far we have produced a hierarchic zonation independently of the taxonomy, consisting of several partitions with fewer and fewer segments, the final one constituting the baseline to object classification. Now the goal is to produce a new hierarchic zonation in which zones (the cells of the zonation) are distributed and nested in accordance to the taxonomy, so that at each level of the new zonation, A) no zone has an adjoining zone located in the same cell of the corresponding level of the taxonomy, i.e. the neighbours of each polygon belong to a different class; and B) parent zones are located in the parent cells of the cells of the taxonomy where their daughter zones where located. In consequence, what we have to do in order to transform the baseline partition into the 0-level partition of the object-tier, is to merge neighbouring granules belonging to the same class of the 0-level (the most detailed one) of the taxonomy.

Hence the previous step to classification is the mapping ***granule_attributes***, which assigns to each granule a value in each of the (surrogate) attributes used for classification. Some of the attributes may consist of the mean (and eventually the variance) in each band of the original image of pixels inside the granule, but many others will relate to non-radiometric information,

[1] *Entia non sunt multiplicanda praeter necessitatem*. A philosophical statement (popularly known as 'Occam's razor') attributed to William of Ockham (1285-1349), that gives precedence to simplicity.

some of them derived from the baseline partition (size, shape and orientation), and some from another ancillary data like e.g. a DEM (mean altitude, slope, aspect, proximity or inclusion of some hydrological feature, like a river or a lake, etc). Also, a surrogate measure of heterogeneity could be given by the number and type of blobs within each granule. For example, blobs could be grouped according to their location in the data space with some clustering algorithm, and then it could be reported both the most frequent spectral class found in the granule and the areal proportion of blobs belonging to a different spectral class. Finally, some attributes may refer to historic data, such as the type of objects from a previously classified partition (e.g. the map that is being updated) that overlap with the granule. The latter, rather than proper attributes, act as links to the attributes of the older objects for comparison. Similarly an adjacency table would serve as a link for comparison with neighbouring granules.

The last operation, probably the most difficult and time consuming (because of the likely human interaction requirement), is the mapping ***classify_granules***. It demands the previous estimation of the range of typical values in each attribute for each landcover class, therefore some ground data must be available beforehand. This mapping can be designed in many ways, but here a general procedure, based upon the ELECTRE[1] method (Roy 1991), will be proposed and briefly outlined. For each granule *i* the actual value of the set of attributes are compared with the typical ones of each class *c*. Then the set of attributes is divided into three subsets according to whether they are concordant, discordant or indifferent to the predicate "granule *i* belongs to class *c*". Indifference arises when there is a range of values within which the attribute can be either concordant or discordant, or when the data is missing or is incommensurable with the class in question (some attributes may be useful to discriminate between some classes but not applicable for some others). There can be also for some attributes a *veto* threshold that precludes the allocation to *c* no matter how many other criteria are concordant. Then a likelihood index is computed for each class according to this division, and the few ones with the highest value are temporarily allocated to that granule as candidate classes.

In a second loop the candidate classes are compared to the one of neighbours in present (and past) partition(s), and as result of the comparison some class candidates are discarded while others get enhanced their likelihood. This routine is stopped when most of the granules show

[1] A family of methods aimed to rank alternatives as an aid to multicriteria decision making.

a candidate class that outranks by far the others. After this, granules not yet classified could be checked on screen by an interpreter, who would either classify manually them or mark them for field inspection if she is unsure about the label. In the field survey, not only the marked granules will be inspected, but also a number of randomly chosen already classified granules, that will serve as samples for accuracy assessment. Once the baseline partition is classified and ground checked, adjacent granules having the same label are merged into a single mapping unit, or polygon. The output of this process is a hierarchic zonation fulfilling conditions A and B stated above, i.e. the final version of the object tier.

In the case of a completely new compilation (in an unexplored territory) of a map of the like of MFE, the lack of formal classes would not allow this kind of approach. Accordingly, either MFE rules are condensed into some reduced set of predefined classes or most of the granules would have to be inspected from the ground. In the second case, an extensive ground survey could be avoided if the imagery is (spatially and spectrally) fine enough to allow species identification. But if this not the situation and the budgetary constraints recommend inspection areas larger than the granules, a higher level of aggregation should be achieved from the imagery. This could be done either manually, by selecting and deleting the arcs separating the granules to be merged, or using e.g. a hierarchical clustering algorithm. The first case would be equal to a computer-aided photointerpretation, where the interpreter has only to click on the arcs to be deleted. Apart from being faster than digitising polygon boundaries on the screen, it would help to reduce the subjectivity of line drawing, since the arcs are already delineated. In the automated method, granules could be assigned to clusters in the attribute space using an overall similarity measure. Geographically adjacent granules belonging to the same cluster would be merged. The similarity threshold beyond which granules are merged would have to be established in a trial and error fashion. The clustering could also be used to stratify the field survey in order to reduce the sampling effort. After the survey, neighbouring mergers having the same label would subsequently be aggregated into the final polygons.

In an scenario where the map has been already compiled and the aim is validation or updating, granules could be used to detect incongruent or changed areas, providing the imagery used to produce the map is comparable to the more recent one (i.e., same sensor and season). Each granule would be compared with the ones overlapping with it in an older partition (and with coetaneous granules belonging to the same polygon, if the aim is validation). As a result, new granules showing an anomalous appearance would be marked for inspection in cases where the nature of the change cannot be inferred from the images. After change identification, each incongruent/changed granule will be either a) included in the polygon where it currently is (if the change/incongruence does not deserve a new label); b) included in a neighbouring polygon (if e.g. there was a mistake); or c) converted into a new polygon, alone or together with other neighbouring incongruent granules. After this, the final shape of polygons is determined by the outer edges of border granules (those which have neighbours belonging to another polygon).

Considering the forthcoming spread of the use of navigation systems, field verification could be entrusted to forestry officers or other field personnel working nearby the place where the inspection is needed. Another future alternative is to carry it out by an unmanned aerial vehicle (UAV) with some imaging instrument. Several companies are developing UAV Remote Sensing prototypes (see e.g. http://www.freewing.com) as a low cost source of 'ground truth'[1] and to perform other specific surveys. UAV RS aircraft carry imaging, positioning, and stabilizing instruments that are operated from a PC-based ground control station that provides command and control of aircraft and payloads, and complete mission planning functions. Envisioning a mid-term scenario where these platforms are commercially available and affordable by regional environmental agencies, ultra-high resolution images, enabling visual identification of species or other features, could be acquired over anomalous granules. Then UAVs would enhance greatly the capabilities of monitoring systems.

Figure 2-4 The NASA/Freewing Scorpion UAV

[1] A RS jargon used to refer to either ground data, high quality existing maps, or higher resolution images enabling a clear identification of landcover types in either the whole scene or in selected disjoint areas.

2.6.3. R-model summary

The outset of the R-model is a scenario where both the budget and personnel of the mapping project are limited. Since it is not possible to retrieve the exact geographic fields with the actual values of the biophysical attributes defining landcover classes over the whole territory, a set of surrogate fields is used instead, consisting typically of (preferably atmospheric and illumination corrected) a Remote Sensing multiband orthoimage, where each field is a band. These fields are also supervenient on the sessile individuals (like e.g. trees and buildings) that compound the commonsense tier, but the correspondence between the value of those fields at a given geographic point and the one of landcover attributes is variable, difficult to determine, and dependent on the spatio-temporal scale of observation.

The R-model assumes that these shortcomings can be tackled if, instead of focusing on point-wise observations, the overall spatial variation of the surrogate fields is studied. The basic premise of the approach is that the spatial variation of the latter coincides with the one of the attribute fields. This constitutes the ***coincidence hypothesis***, which links the R-model to the I-model at the field-tier. At the object tier, what the coincidence hypothesis states is that blobs correspond to patches. In other words, it is assumed that the structural-functional units of the image correspond to structural functional units in the landscape.

The R-model, as well as the I-model, assumes that patchiness (granularity) is ubiquitous and can be analysed at different levels of generalisation (grain), in accordance to the hierarchical structure of the landscape (from individual plants to biosphere). They set their basic level at patches defined as a contiguous area of similar dominant species, physiognomy and physiography, so that patches of an inferior level, nested into the formers, are traced over. The question now is to what level of the patch hierarchy the blobs of an image belong.

The answer relies on the fact that the size of patches of the so defined basic level varies widely, from a few thousands sq meters to hundreds of sq kilometres. Since it is very unlikely that all the blobs of an image reach the size of the patches of interest, most blobs will have to be aggregated into bigger units, providing we are working on high resolution imagery. A new commitment is made, called the ***size hypothesis***: blobs under a certain size correspond to patches of no significance for users, and therefore need not to be retained as individual entities. Conversely, segments exceeding this size are big enough to have a meaning for users and thus may potentially become mapping units after classification. These segments constitute the *minimal semantic units* of the R-model, and are termed ***granules*** to stress they are seen as if they had no further parts, i.e. they are the 'atoms' of geographic reality in the object-tier. Thus granules are the minimal cells of a *zonation* where their minimum admissible size is normally equated to the MMU. This zonation is called ***baseline partition***, to indicate that it constitutes the starting point of an object-based analysis from which the final configuration of the object-tier will be derived.

The strategy proposed to derive the baseline partition is to construct, based upon the multiband image, a hierarchic zonation whose top layer is the searched partition. The first layer corresponds to the partition defined by the morphology of the image, i.e. a primal sketch with the contour of the basins of attraction (blobs). In order to derive the morphology, the image is considered as a dynamic system of nodes (pixels) that interact strongly or weakly depending on their radiometric similarity. This system evolves to a steady state where the singular nodes (those which do not interact with all their neighbours) define the morphology. The resulting regions (blobs) are subsequently aggregated via a region-merging algorithm that merges iteratively (only one merge per segment and iteration) blobs to their most similar neighbour until all the segments are larger than MMU. The region merging stage is based on the ***correspondence hypothesis***, which assumes that if two adjacent segments are radiometrically similar, they will be semantically similar too.

The baseline partition is finally transformed to create, based upon the classification scheme of the map, the final zonation, consisting of polygons representing the geographic objects of interest. The first step in this process is to compile the set of surrogate attributes upon which the granules will be classified. Once the baseline partition is classified and ground checked, adjacent granules having the same label are merged into a single mapping unit, or ***polygon***. The output of this process is a (hierarchic) zonation associated to a (hierarchic) taxonomy, i.e. the final landcover map.

It is worth noting that the R-model works with <u>four types of areal units</u>, each one supervenient on the former. The <u>data basic unit</u> is the *pixel*. Pixels are artificial units whose value has been measured from the *support* region (the GIFOV) of the detectors. The <u>image basic unit</u> is the *blob*. Blobs are 'natural' units, in the sense they conform structural (their parts –pixels- are more similar between them than to their surroundings) and even functional (under the dynamical network analogy, their parts –nodes-interact more between them than with their surroundings) wholes. The process of blob detection is completely data-driven, there is no imposition from the user. The semantic basic unit is the *granule*. The main difference between blobs and granules is that *all* the granules may be categorised as instances of some class, whereas *not all* the blobs may constitute a big enough entity as to be considered an object of its own. The minimum size of granules determines the level of spatial generalisation applied to the territory. Finally, the conceptualisation applied to geographic reality is given by the classification scheme of the map, which is reflected in the way granules are grouped into classified objects: the *polygons* or mapping basic units. Polygons are aggregations of neighbouring granules that have the same meaning under a particular view of geographic reality: they foreground the geographic objects we want to see in the territory. The classification process is dependent on granule attributes, whose definition and value distribution for each class relies upon the user. Different decisions during this process will yield different models of the territory.

2.6.4. Further considerations on the R-model

The construction of the R-model is a journey from complexity to simplicity, at the expense of increasing uncertainty, where the successive representations are simpler (have less regions) than the previous ones. The first representation of the territory, the most detailed one, is the RS orthoimage. It already contains a great deal of generalisation, proportional to the pixel size. In this sense, the regularisation effected by the sensor yields a model of the imaged scene. This model is isomorphic to the visible territory, i.e. its structure bears an exact correspondence with the distribution of ground features observable at that scale, preserving their topological relations. The spatial structure of the image is given by the luminance (and chrominance) variations that take place across it. Such structure is formalised in the next model, which makes explicit the salient parts (the boundaries of the basins of stable attractors) of it. As a result we move from a huge amount of artificial regions (pixels) to a fewer number of 'natural' regions (*blobs*). The relation between the new model and the territory is grounded on the *coincidence hypothesis* between blobs and patches.

The next step in the model build-up is to acknowledge that the only way to tackle the hierarchical granularity of the territory is to set a level of generalisation below which we do not care *what* and *where* the parts are. In doing this we assume the *size hypothesis*: that patches (and therefore blobs) below a certain size will not deserve representation. The actual threshold chosen is arbitrary, and it will determine, together with the level of abstraction of the classification scheme, the validity of the size hypothesis. Once defined, blobs have to be merged until they reach the chosen size. The aggregation relies on the (local) *correspondence hypothesis* between radiometric (spectral) similarity and semantic (taxonomic) similarity. The output is a representation of the territory consisting of less regions (*granules*), which is the baseline to the final stage.

The final model represents a conceptualisation of the territory that consists of objects (*polygons*) inserted into categories or classes. Each class-concept consists of a name or short predicate that can be defined by either *intension* or *extension*. The *intensional* definition of a landcover type comprises the properties and relations a landcover patch needs to have in order to belong to the class. The extensional one is the set of all the patches that befit the class-concept. The intension has to be translated into a set of surrogate attributes more related to the appearance of granules in the image than to the actual properties of the biophysical cover of the terrain onto which the granules project. If the translation is correct, patches that correspond to granules showing the typical values of a given class, will have this class-concept as a true proposition. Therefore they will constitute the *extension* of this landcover type on that territory. The extensional definition is condensed to form the final representation, by retaining only maximal sets of connected granules of the same class, i.e. polygons.

The accuracy of this representation (i.e. whether boundaries are placed on the right location and whether granules have been correctly classified) relies, on the one hand, on the significance of the surrogate attributes as class marks and on the appropriateness of the chosen classifier, and on the other, on the validity of the three hypotheses employed to derive the baseline partition. The first issue is beyond the scope of this thesis. The second has already been discussed, but it is worth noting that it is inseparable from the former. Mapping errors are "forcible deviations between a representation and actual circumstances" (Chrisman 1991). These circumstances have to be defined with regards to the conceptualisation of the territory given by the map legend. Therefore the accuracy of the baseline partition can only be assessed once the granules are classified.

However, the baseline partition can be evaluated as a model of the image structure. Given a target level of generalisation (i.e. the MMU size), the best partition would be the one that minimises deviations between the value predicted by the model (the mean of the granule) and the actual data value (the DN of pixels belonging to that granule). This idea will be used in the next chapter to assess the filtering scheme.

To finish up, some advantages of using granules instead of individual pixels as the basic units to classify landcover should be stressed. **First**, granules are enclosed by boundaries derived from genuine discontinuities, whereas pixels have an artificial shape that keeps no relation with the spatial distribution of the geographical phenomenon under study. **Second**, as a result of the former, granules are less sensitive to the MAUP than pixels. This means that it can be expected that, given a target MMU size, the baseline partition is almost invariant to the pixel size of the input image, providing i) the latter is in the range discussed in last paragraph of 2.6.2.1.4; and ii) the procedure includes a line generalization algorithm that handles suitably the fractal nature of boundaries. **Third**, using granules instead of individual pixels as basic units reduces considerably the computational burden of the analysis. **Four**, new attributes related to shape, topology (context) and heterogeneity can be derived for granules that cannot for pixels. **Five**, granules are compatible with an object-oriented approach based upon the hierarchical patch model, i.e every granule may become an instance of a type of patch of the hierarchic level of interest, while no single pixel or even small blob can become a mapping unit of its own. And **six**, classified granules are best suited for monitoring change than classified pixels, since the former have an intrinsic meaning (and hence an identity) that the latter lack.

2.7. Class-concepts and the inadequacy of the spectrometric approach

Be it the I-model or the R-model, the last tier represents a set of geographic objects, each one having a different meaning than that of its neighbours. The identified objects are located in the cells of a *zonation* projecting onto the territory, where at the same time these cells are located in cells of a *taxonomy*. The latter consists of the set of landcover types included in the classification scheme of the map. Each landcover type is a class-concept that can be defined by either intension (the set of properties distinguishing it from all others) or extension (the set of polygons belonging to it). In this section I discuss the implications of the nature of class-concepts on the appropriateness of using the spectrometric approach to classify RS images.

The spectrometric approach (SMA) conceives the territory as made of distinct homogeneous materials (landcover classes) that are spatially distributed into pieces larger than a pixel. Each class may have several typical spectral signatures, but all of them are supposed to be distinguishable from the ones of other classes. The set of typical signatures is used to construct the intensional definition of each class. In doing so, the SMA assumes that class signatures are separable (not similar between instances from different classes). The signatures are recorded by remote sensors and subsequently classified by drawing the boundaries of the region(s) of the data space occupied by each class. Signatures inside that region constitute the extensional definition of the class. Under the SMA, it is expected that the projection of this extension onto the territory coincides with the one that would have been obtained should the proper intension (related to biophysical features) be applied to exhaustive ground observations.

The first difficulty arises from what is considered an instance. Under SMA, class-concepts are *mass nouns* (like e.g. water) that refer to types of homogeneous materials. Under such view, any arbitrarily delimited piece of terrain made of this material is still a referent to that noun. Therefore an instance can be any piece of terrain (represented by e.g. a pixel), provided it is large enough as to include the features constituting the biophysical intension of the class. This is in sharp contrast to the entity view of the territory portrayed in the map, where instances consist in polygons. The latter are better conceived as referents to *count nouns* (like e.g. lake), i.e. as instances of types of geographic objects.

Both views would not be in conflict if all the 'mass' instances that can be individuated in a given polygon conform to the biophysical intension of the class, as in a typical water/lake example. However, polygons usually include some parts that do not comply with its intensional definition (e.g. a lake with some small islands in it). Such parts have a reduced extension (smaller than the MMU size) that precludes their inclusion in the map as separate units. Thus if the MMU size, as it is often the case, is several times larger than the pixel size, the geographic extension of classes will differ considerably if mass-instances (pixels) are used instead of count-instances (granules). In this sense, the 'entity' model and the 'materials' model are incompatible.

In 1.11, the inconsistency of SMA when instances are regularly distributed was related to either pixels smaller than the classification disk or to an excessive amount of *mixed pixels*. If in order to tackle it, granules are used instead of pixels, the situation is neither improved. Granules may include inhomogeneities too, i.e. gaps from a different class. Therefore, the mean signature obtained from pixels inside the granule will be corrupted by such inhomogeneities. This means that, even granting the existence of separable class signatures, it cannot be expected that separability is preserved when granules are taken as class instances.

Then, if the SMA cannot work properly with either pixels or granules, the only alternative consists in using *blobs* as instances, providing the pixel size is close to the classification disk. The pixels of a blob are radiometrically similar by definition. Then, given their proximity, they can be expected to be semantically similar as well (recall the correspondence hypothesis). If there are separable class signatures, then blobs are the best candidate areas from which individual signatures are to be extracted. Having defined what proper instances should be, we are ready to discuss the next, more difficult problem.

The SMA assumes that the radiometric intension (hereafter r-intension) of classes yields roughly the same geographic extension than the biophysical intension (b-intension). This basic assumption requires that **i)** regions of geographic space conforming to the b-intension of classes are mutually disjoint (e.g. no patch can be 'agriculture' and 'forest' simultaneously); **ii)** regions of data space conforming to the r-intension of classes are mutually disjoint (e.g. no signature can be 'agriculture' and 'forest' simultaneously); and **iii)** signatures of regions of geographic space conforming to the b-intension of a given class conform only to the r-intension of that class (e.g. if a patch is 'forest' biophysically, then the signature of the corresponding blob is 'forest' as well). Note that all the signatures are now 'pure', since in principle there are no 'mixed' blobs. Therefore, if the previous assumption is valid, class overlap in the data space will be insignificant, meaning that there is a negligible proportion of signatures with 'mixed' neighbourhood (see 1.11).

Note also that signatures of a given class may differ considerably to one another as long as none of them is similar to signatures of another class. However, similarity is a comparative term requiring some consistency on the side of class signatures. If we equate similarity to proximity in the data space, the former condition means that each individual signature must be closer on average to signatures of its own class than to signatures of other classes. In other words, the nearest neighbours of any signature should belong to the same class, i.e. it should posses a 'pure' neighbourhood in the data space. This situation can only occur when the boundaries between regions of different class are constituted by quasi-empty space.

Assuming the existence of natural boundaries in the data space, the former discussion can be restated as follows. The extension of each class in the data space is a cluster or a group of clusters, where each cluster is a more or less compact 'cloud' of signatures surrounded by empty space. Note that a cluster may be nested onto a larger cluster of different class (like an air bubble trapped inside a glass ball) as long as it is isolated by empty space. Consider now a single cluster and the signatures populating it. The SMA assumes that most patches from which these signatures were collected conform to the biophysical intension of that class. If there is some signature within that cluster corresponding to a patch that conforms to a different biophysical intension, it will be misclassified in all likelihood. In short, in order for the SMA to be successful, there should be a prevailing class in each cluster of the data space.

Such nice correspondence is only possible if class-concepts are also natural groupings of natural entities. If class-concepts were artificial, their intension could be modified arbitrarily as to make them fit a particular purpose. We could e.g. lower the TCF threshold that separates forest from sparse woodland from 35% to 5%. The new intension would cause a modification of the extension, so that some patches would change class. If the corresponding signatures were located in clusters where some signatures still belong to the old class, the clusters involved would consists of a mixture of classes, making invalid the SMA assumptions. So the key point is to assess to what degree class-concepts are natural.

Unfortunately, the weight of evidence suggests that class-concepts are less 'natural' than we assume. If class-concepts were natural, their extension should consists of groups of patches closely similar to one another and clearly different from patches of other classes. However, we saw in 2.2.3 that instances are aggregated to a class by means of a loosely defined family resemblance. For example, despite an orange tree plantation is biophysically more similar to an artificial forest than to a wheat field, it is usually classified as 'agriculture'. Also, patches whose biophysical description falls in between the intension two classes would be rare. But in the same way there are no infinitely thin lines in geographic space, there are no clear-cut distinctions in 'categorical space'. As a point of fact, it is not difficult to find instances of intermediate types between any given pair of classes. Even the most natural of class-concepts, the *species*, has boundaries that can be blurred. Vg. many plant species interbreed, and the resulting hybrids are fertile.

Classes are always human creations, in the sense we impose the boundaries separating them. What we actually classify is not patches, but conceptions of patches based upon observations. And in the multiplicity of observable properties of each of these patches, every observation is contingent on the observer's selection of the observed (Whittaker 1962). Hence, the way in which we group entities into classes does not necessarily reflect a natural order of things, but rather the subjective judgments of mapmakers and taxonomists. Perhaps the most conspicuous evidence of the latter is our present position in the taxonomy of the animal kingdom[1]. However, note that 'subjective' does not imply 'arbitrary'.

From the plenitude of boundaries that can be identified in a data set, the analyst chooses those that he/she subjectively judges as relevant for modelling the phenomenon under study. This decision is ultimately grounded on quantitative differences, therefore there is always some objective basis. Then, in the continuum from natural to arbitrary, from fact to artefact, the actual position of the output model would depend on the level of consensus about its usefulness that can be reached among its users after having used it. The results from models 'fitting nature' in the opinion of users will be treated as facts, while the results from not so successful models will be catalogued as artefacts. As a consequence unsuccessful models will be discarded in favour of more useful ones. This 'natural' selection is an evolutionary process guiding scientific progress and not only. As a point of fact, human vision is the best modelling system of optic data known so far.

In short, there are many ways in which fiat geographic objects can be conceptualised and individuated. Hence it is too optimistic to expect that there is always a bijection between radiometric and biophysical intensions, i.e. that the regions delimited in the data space keep a one-to-one correspondence to the geographic concepts that we use to divide up the landscape into meaningful chunks. On the one hand, classes are formed by family resemblances following a set of typical instances. For any given class, it is hard to find a character that is shared by all its members and that is absent in members of other classes. Therefore the

[1] In his first edition (1735) of *Sistema Naturae,* Carl Linnaeus, being unable to find a generic character distinguishing man from ape, put them in the same *genus* (respectively *Homo diurnus* and *Homo nocturnus*). After having received strong criticism and fearing religious prosecution, he split them onto separate genera in a later edition (1748). A quarter of millennium later, and having stronger biomolecular evidences on our close relationship with chimpanzees and other apes, taxonomists do not yet dare to undo Linneaus' correction.

mathematical definition of classes is an ill-posed problem[1]. On the other hand, the surroundings (and even gaps in their interior, 2.2.7) of geographic objects may be significant for classification. Therefore a sound classification cannot rely solely on radiometric signatures. In addition there is a wealth of relational features that should be taken into account (see 2.6.3).

The conclusion is that, rather than in its truthfulness, the validity of a given thematic map relies on its usefulness for a particular purpose. The latter determines the type of information (model) about the territory that is required. In most applications, the aim is to produce an entity model of the territory in which the regions individuated are instances of some discrete abstractions (concepts of geographic objects). In such cases, an object-oriented approach (compatible with the landscape conception of the target model, i.e. the hierarchical patch model) to classification is more appropriate than the spectrometric approach (which uses a quite different conception –the piecewise homogeneous model- of the landscape). This chapter has tried to set forth the conceptual principles underlying such approach.

[1] Maybe a consequence of Gödel's (Gödel 1931) theorems, that could be restated as follows in the context of landcover mapping:

i) There are configurations of landcover that cannot be assessed by a non-contradictory set of rules and therefore cannot be classified by a non-contradictory scheme.

ii) It is impossible to prove that the classification scheme is non-contradictory using its set of rules.

CHAPTER 3

The baseline method

A new theory is always announced with application to some concrete range of natural phenomena; without them it would not be even a candidate for acceptance.
Thomas S. Khun (1962), The structure of Scientific Revolutions

3.1. Introduction

In this chapter, the conceptual framework set forth in the previous chapter is used to construct a method that enables the transformation of a numerical raster representation of the territory (the input RS image) into a symbolic shape-oriented representation (the vector layer of the baseline partition). The latter is the starting point of an object-oriented classification in which each individual region (granule) is evaluated for inclusion as a member of one of the user-defined object classes. Hence the name of baseline partition, and thus the procedure to obtain it is hereafter called the *baseline method*.

Before proceeding, it is worth stressing again that objects are class instances, and in the case of landcover they are patches that qualify as referents of the concepts we use to divide up the landscape into landcover types. Then, in order for a patch to become a geographic object, it should be such that it can be *conceived* as a unitary whole. This implies that it should be perceived as internally coherent and externally differing from its surroundings, and not least, it should have enough extension as to deserve representation as an individual entity. The latter is defined through the MMU size, so that patches below this size can never become instances of classes at that level of generalisation. Therefore, under the object-oriented approach, it is required that the spatial units subject to classification exceed the MMU size. The granules of the baseline partition provide such units. Note that the previous assertion does not preclude the inclusion in the analysis of some attributes derived from the distribution and characteristics of some sub-units (as blobs) within each granule.

Also, some caveats must be stated now. The method presented here to derive the baseline partition is by no means the only possible way to achieve it. It is simply one that follows naturally from the realistic model previously expounded. There are several somehow arbitrary choices that may be changed in subsequent versions, namely: i) the filter type and intensity; ii) the dissimilarity measure used to compute the gradient magnitude image and the radiometric distance between segments; iii) the preliminary simplification of the watershed partition (currently discarded); iv) the region merging algorithm used; and v) the line generalisation algorithm applied in the vectorisation of the partition. Any of these choices could be changed without altering significantly the scheme.

The current implementation of the method (written in IDL[1]) is only intended to show the type of results than can be achieved in real images by applying the concepts presented in this thesis. It is still far from optimality, both from the computational and design points of view. The path to be followed in order to have it fully operational is too long to be completed within this thesis. Therefore no effort is made to assess thoroughly its performance. As a matter of fact, the only evaluation that has been made was aimed at choosing a convenient level of diffusivity to filter the images.

3.2. Method overview

The baseline method is an implementation of the procedure described in the previous chapter to derive an uncommitted model of the territory (i.e. a zonation with no associated taxonomy) that is the base for the construction of the object tier of the R-model. A description of the equivalences between the operations defined in Chapter 3 and the algorithms used in the current implementation of the baseline method is shown in the table below:

R-model operation	**Description**	Baseline method algorithm
steady_state	Transforms the input image into an almost piecewise constant image	GIWEPS
get_morphology	Computes the gradient magnitude of the former image, and locate the pixels bounding the area of influence of each gradient minimum	Gradient Watershed
label_blobs	assigns a unique numeric label to each area of influence (blob)	Watershed
build_granules	Aggregates blobs into segments larger than the MMU size (granules) and convert the output into a vector layer	SCRM vectorise

Table 3-1. Equivalences between R-model operations and baseline method algorithms

The sequence is the following. The input image (previously ortho-rectified to some planar cartographic projection, and eventually resampled to a suitable pixel size) is filtered in order to get rid of superfluous gradient minima created by texture and/or noise. The output of the

[1] Interactive Data Language (www.rsinc.com). An application-specific programming language specially suited for processing big multidimensional arrays like RS images.

filter (Gradient Inverse Weighted Edge Preserving Smoothing, GIWEPS) is an almost piecewise constant image, in which each uniform region is the area of influence of a gradient minimum. The gradient magnitude of the dissimilarity measure (Euclidean distance, for single band images, or the Normalised Vector Distance, NVD, for multiband images) is computed, and the output image is searched for local minima. The area of influence of each minimum is contoured and labelled with the watershed algorithm, and then the resulting regions are merged through a region-merging algorithm (Size Constrained Region Merging, SCRM) until they all exceed the minimum mapping unit (MMU) size. SCRM uses the original data to compute the dissimilarity measure between segments. Finally, the labelled image with the baseline partition is later converted into a vector. Granule attributes could be optionally compiled at this last stage. The workflow of the baseline method is shown in figure 3-1.

The remain of this chapter is as follows. The procedure used to evaluate quantitatively the method is explained (3.3), and so is the dissimilarity measure (3.4). Then computation of the gradient magnitude image is briefly described (3.5). Although the filtering (3.6) precedes the latter in the workflow, it is introduced after the gradient because the latter should be clearly understood first. Then watershed transform, which is the cornerstone of the method, is explained in detail (3.7). Afterwards, a novel region merging (SCRM, 3.8) is presented that lead to the baseline partition. To end up the methodological part, the method to convert the latter into a vector layer is briefly described (3.9). Finally, some examples of the results obtained with the method in some RS images are shown (3.10) and discussed (3.11).

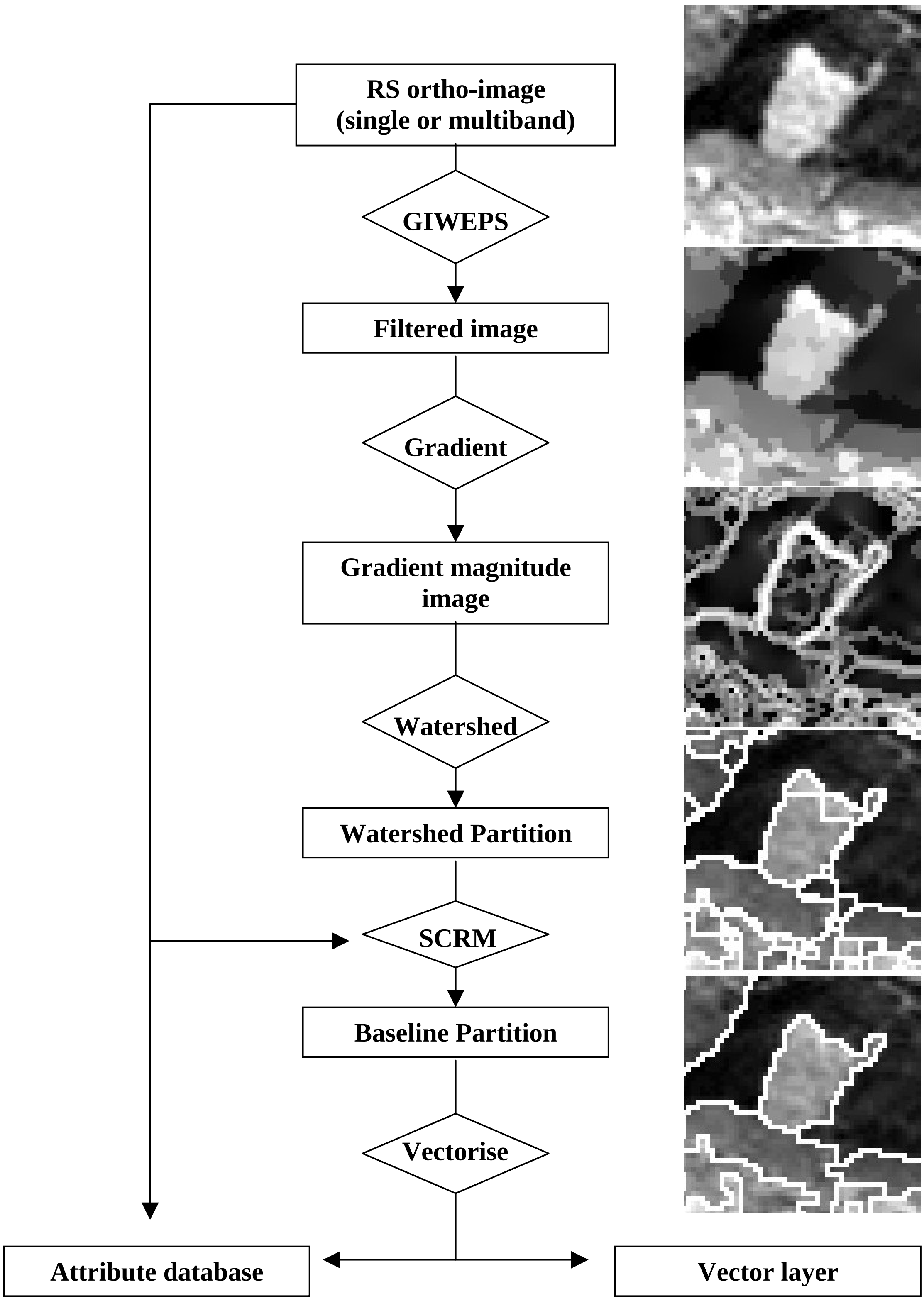


Figure 3-1. Baseline method workflow

3.3. Method evaluation

The baseline partition is a model of the territory that is constructed by two separate mechanisms of generalization: the regularisation effected by the sensor and the aggregation of primal regions (the areas of influence of gradient minima) according to their radiometric similarity. The segmentation method used to implement the second mechanism should ideally be uncommitted, i.e. it should need no *a priori* knowledge from the user, so that the only input required is the MMU size. Then, the overall segmentation process can be seen as an automated zone design problem (Openshaw 1978) that consists in aggregating n regions (pixels) from a primary model (the image) into k larger regions exceeding the MMU size. The number k of regions is an indicator of the size or scale of the target model, i.e. k reflects the level of generalisation applied to the imaged territory.

Since later on granules are considered semantically homogeneous, the baseline partition can be evaluated under the piecewise homogeneous model underlying the spectrometric approach, just by equating semantic homogeneity to radiometric homogeneity. With this analogy, the baseline partition can be viewed as a model of the image that is piecewise constant, each piece being a granule. Once the target size of the model is chosen, the best model would be one that minimizes at each location (pixel) the deviations between the value predicted by the model (the mean of the granule to which it belongs) and the actual value of that pixel. An overall measure of the goodness of fit of models is the root mean square error (RMSE). RMSE of models increases as their size decreases. When the model size is maximal (i.e. when each pixels is a region), RMSE is zero, but actually there is no model but the data itself. When the model size is minimal (the whole image is the only region in the model), the RMSE is equal to the standard deviation of the image. Therefore a normalized RMSE (hereafter nRMSE) can be obtained by just dividing it by the standard deviation of the image. Then the evaluation of alternative methods for deriving the baseline partition can be based upon the nRMSE, which is given by the following expression:

$$nRMSE = \sqrt{\left(1/(n \cdot l)\right) \cdot \sum_{b=1}^{b=l} \left(1/\sigma_b{}^2\right) \cdot \sum_{i=1}^{i=n} \left(x_{ib} - \bar{x}_{ib}\right)^2} \qquad (3.3.1)$$

where *n* and *l* are respectively the number of pixels and the number of bands of the image, σ^2_b is the variance of band *b*, x_{ib} is the value of pixel *i* in band *b* and $\bar{x}_{ib}$ is the mean in band *b* of the granule to which pixel *i* belongs. nRMSE will be used to select suitable values of the filter diffusivity. Note however that the analogy, upon which this error measure is based, is dubious from the object-oriented approach perspective, for the latter admits explictly the presence of inhomogeneities within each object. Therefore, the nRMSE should be taken only as a preliminary approximation that is only suitable for the uncommited stages of the modelling process.

3.4 The normalized vector distance (NVD)

In the case of multiband images, a dissimilarity measure between signatures has to be defined that provides a metric to the multidimensional data space. The most straightforward alternative would simply be the Euclidean distance between points in the data space. However, it does not model properly chromatic differences, since the latter are, at least from a perceptual point of view, not isometric to Euclidean distance, due to the differential sensitivity of the human eye to different wavelengths and luminance (Kaiser & Boynton 1996). If the angular difference between signatures is used instead, hue discrimination would be considerably improved, but in turn luminance differences would not have been taken into account at all (Wesolkowski 1999). Hence, if the goal is to imitate photointerpretation results, the best measure would be one that combines both luminance and chromatic contrasts. Baraldi and Parmiggiani's (Baraldi & Parmiggiani 1996) Normalised Vector Distance (NVD) fulfils this requirement.

Let $X(x_1, x_2, ..., x_k)$ and $Y(y_1, y_2, ..., y_k)$ be two signatures from an image with *k* bands. Let *L*(X,Y) be a normalized measure of the luminance contrast between both signatures. *L* is defined by the expression:

$$L(X,Y) = \min\{|Y|/|X| \, , |X|/|Y|\} \qquad (3.4.1)$$

where |X| and |Y| are respectively the moduli of X and Y. *L* ranges from 0 to 1. Let now *C*(X,Y) be a normalized measure of the chromatic contrast between X and Y. *C* is defined by:

$$C(X,Y) = (\pi/2 - \alpha)/ \pi/2 \qquad (3.4.2)$$

where α is the angle, expressed in radians, between the vectors X and Y, that varies from 0 to $\pi/2$, providing the dynamic range of the image is positive. This angle is given by the formula:

$$\alpha = \arccos((X \circ Y) / (|X| \cdot |Y|)) \qquad (3.4.3)$$

where the symbol $\circ$ is the scalar product ($x_1 \cdot y_1 + ... + x_k \cdot y_k$). *C* also ranges from 0 to 1. Then the Normalised Vector Distance is defined as the complementary of the product of both measures:

$$NVD(X,Y) = 1 - L(X,Y) \cdot C(X,Y) \qquad (3.4.4)$$

Note that an additive combination (e.g. NVD=1-(*L*+*C*)/2) is not suitable since, when one of the two measures is low, NVD should be high no matter how high the other measure is. NVD defines a metric on the data space *D* (Baraldi & Parmiggiani 1996):

1) $NVD: DxD \rightarrow [0,1] \in \Re^+$
2) $NVD(X,X) = 0 \quad \forall X \in D$
3) $NVD(X,Y) = NVD(Y,X) \quad \forall X,Y \in D$
4) $NVD(X,Y) \leq NVD(X,Z) + NVD(Y,Z) \quad \forall X,Y,Z \in D$

NVD was conceived as a dissimilarity measure for segmentation algorithms. In this thesis it is also used to compute a surrogate gradient magnitude from multiband images, and to perform an edge preserving smoothing on them, as it is shown in the next two sections.

3.5. Gradient magnitude image

In the previous chapter (2.6.2.2), the stable attractors of a geographic field (RS image) were assimilated to points of minimum variation, i.e. pixels where the gradient is locally minimal. The basins of attraction (area of influence) of gradient minima constitute the regions (blobs) of the first partition, hence the need for a gradient magnitude image. At each pixel of the original image, the magnitude of the variation is given by the modulus of the gradient vector,

∇. The modulus |∇| is computed according to the first derivative of the image in two orthogonal directions (horizontal and vertical), which can be approximated as follows:

$$\left|\nabla_{ij}\right| = 100 \cdot \sqrt{\left(x_{i-1,j} - x_{i+1,j}\right)^2 + \left(x_{i,j-1} - x_{i,j+1}\right)^2} \;/\; r \qquad (3.5.1)$$

where $|\nabla_{ij}|$ is the gradient magnitude of the pixel located at column *i* and row *j* of the image. Note the value x_{ij} of the central pixel is not considered in the calculation, only that of adjacent pixels in the North, East, South and West directions. In order to have this magnitude normalised, the output of the square root is multiplied by 100 and divided by the dynamic range *r* (usually 255) of the image.

In the case of multiband images, the following expression is used:

$$\left|\nabla_{ij}\right| = \sqrt{\left(100 \cdot NVD(\vec{x}_{i-1,j}, \vec{x}_{i+1,j})\right)^2 + \left(100 \cdot NVD(\vec{x}_{i,j-1}, \vec{x}_{i,j+1})\right)^2} \qquad (3.5.2)$$

By using NVD, a surrogate gradient magnitude image can be derived that describes the variations in similarity of adjacent pixels across the image. Finally, the local minima of the gradient magnitude image are those pixels whose value is lower than the one of their eight neighbours. In the unusual case of plateaus (regions with equally valued pixels), there are no proper minima, and the centroid of the plateau is selected as a local minimum representing the region. Note that high plateaus (flat regions surrounded of pixels with a lower value) are not possible in gradient magnitude images, since high valued pixels are all located in ridges.

3.6. Image smoothing

Due to the hierarchical patchiness of landscape, RS images of any resolution show texture of varying degrees across all their extension. In addition, there always be some noise produced by the imaging instrument during the acquisition. If the gradient magnitude image is computed without any conditioning of the RS image, the computation will result in an intricate structure full of edges and local minima, especially in areas with coarse texture. Since each segment of the first partition is the area of influence of a gradient minimum, the partition will consist of a myriad of small segments. In watershed partitions, boundary pixels belong to none of the segments they separate, therefore their value is not included in the mean of segments. As the number of boundary pixels is greater than that of inner pixels, the subsequent region merging algorithm will use a small proportion of pixels in the calculation of radiometric distances, making the initial stage of the process unreliable in zones of coarse texture. Therefore the structure of the image has to be simplified in order to get rid of gradient minima produced by texture.

The simplification should be performed without altering the resolution of the image, that is, it should act only upon the unresolved elements of the image that produce texture, leaving untouched the elements corresponding to edges. Such process is commonly called Edge Preserving Smoothing (EPS). EPS is family of filters from which the simplest one is perhaps the median filter, which replaces each pixel with the most frequently found value within a window centred at each pixel; see e.g. (Abramson & Schowengerdt 1993) for a review on EPS. These filters are preferred over conventional smoothing techniques like simple averaging or gaussian filtering because the latter blur edges and thus make them harder to identify. In contrast, EPS filters adapt the process to the structure of the image, so that the local operator (filter) applied to pixels is different at each position. A precursor of the type of scheme used here can be found in (Wang, Vagnucci, & Li 1981).

An interesting property of most EPS algorithms is their convergence to a non-flat image. That is to say that, when the process is applied iteratively, consecutive output images are increasingly similar to each other, up to a point when change is negligible. Output images beyond this point still resemble the original image during very long time (thousands of iterations). In contrast, non-adaptive schemes decay rapidly to a flat image when they are applied iteratively. After convergence, the output image is almost piecewise constant, and each homogeneous region corresponds to the area of influence of a gradient minimum.

The EPS introduced here, hereafter called GIWEPS (Gradient Inverse Weighted Edge Preserving Smoothing), is analogous to and easier to implement than the widely-used Perona-Malik (Perona & Malik 1990) filter. In general, any adaptive filter yielding piecewise constant (segmentation like) results can be used to simplify the gradient magnitude image. Notwithstanding the former, in practice, even linear (non-adaptive) filters could be used, providing the boundary displacement they cause is not relevant for the application at hand, as e.g. when the pixel size is several times smaller than the required positional accuracy of the final map.

Convergence enables the analogy with a physical system that reaches a steady state far from equilibrium. In 2.6.2.1.1, the image was conceived as the initial state of a planar dynamic network consisting of triangular meshes made up of nodes (pixels) connected through links via which the nodes interact during several cycles. This iterative interaction can be seen as non-linear diffusion process that can be formalised by partial differential equations with the original image as initial condition (Weickert 1999). However, a simpler notation will be used here. Let x^t_{ij} be the value of the pixel located at column *i* and row *j* of the image at iteration *t*. The new value x^{t+1}_{ij} is given by the formula:

$$x_{ij}^{t+1} = \sum_{c=-1}^{c=1}\sum_{r=-1}^{r=1} w_{cr}^{t} \cdot x_{i+c,j+c}^{t} \qquad (3.6.1.)$$

where

$$w_{cr}^{t} = g_{cr}^{t} \Big/ \sum_{c=-1}^{c=1}\sum_{r=-1}^{r=1} g_{cr}^{t} \qquad (3.6.2)$$

and

$$g_{cr}^{t} = 1/\left(1+\left|x_{i+c,j+r}^{t} - x_{ij}^{t}\right|^{p}\right) \qquad (3.6.3)$$

In other words, the new value of x_{ij} is the weighted mean of a 3x3 window centred at the pixel, where the weight of each neighbour is inversely proportional to its radiometric distance to x_{ij}. The more similar a neighbour is to x_{ij}, the greater its weight, and vice versa. The power *p* can be any positive integer. The greater the value of *p*, the more similar the output image is to the original one, the more gradient minima remain, the lower the RMSE and the faster the convergence (figure 3-1). Hence p is the parameter determining the decay of diffusivity with radiometric dissimilarity. Note that for *p*=0, the filter would we equal to a simple 3x3 mean filter.

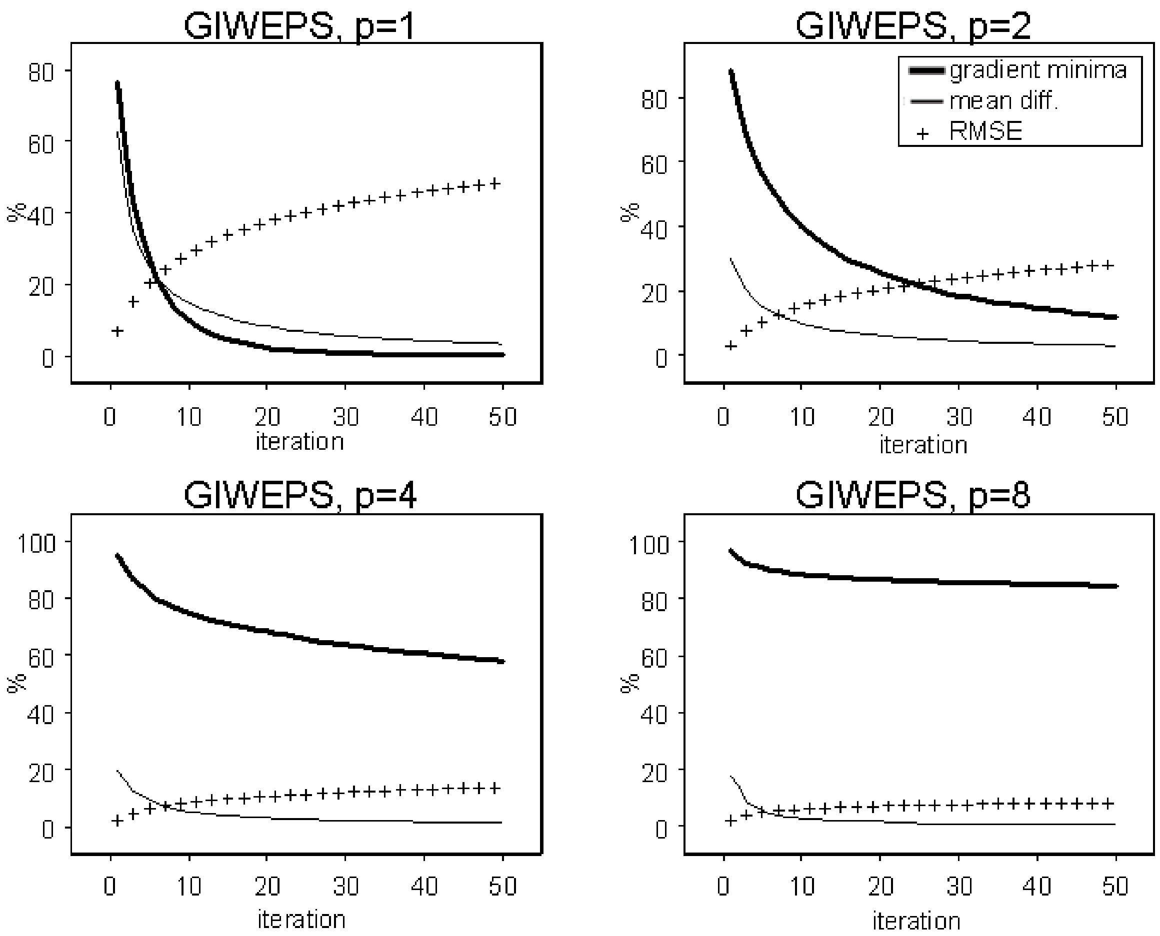
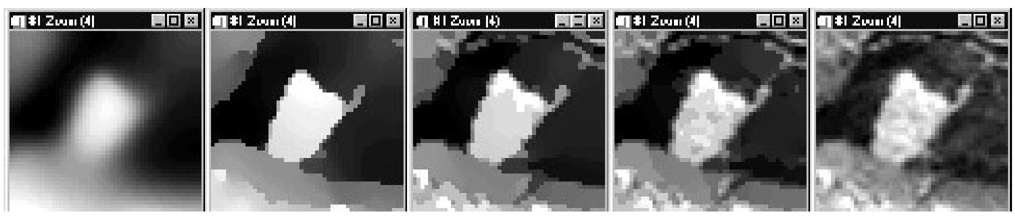


Figure 3-2. The impact of the diffusivity parameter *p* on GIWEPS output. The number of gradient minima remaining after each iteration (thick line) is expressed as a percentage of the one in the original image. The thinner line is the mean absolute difference between consecutive output images, expressed in hundredths of DN. Convergence is reached when this line flattens. The plus line is the RMSE at each iteration of the output image when compared to the original one, expressed as a percent of the standard deviation of the latter. The images below show a 50x50 pixel (1.5 km^2) detail the output images after 50 iterations with different values of p. The last image is the corresponding detail of the image (1a) used for the computation.



In the case of multiband images, the NVD is used as the radiometric distance between each pixel and its neighbours. Hence equation 3.6.3 is modified as follows:

$$g^{t}_{cr} = 1/(1+(100 \cdot NVD(\vec{x}^{\,t}_{i+c,j+r}, \vec{x}^{\,t}_{ij}))^{p}) \qquad (3.6.4)$$

NVD is multiplied by 100 to express it in percentage. Note that although NVD is computed with signatures (vectors) of pixels, the resulting weight is applied separately to each band. Hence the first computation is the NVD between each pixel and its neighbours, and then each band is processed separately.

In both single and multiband images, convergence is theoretically attained when the number of gradient minima hardly decreases in successive iterations, meaning that the remaining attractors are stable (see 2.6.2.1.1). Since this number decreases asymptotically (reaching zero for $t=\infty$), an arbitrary threshold has to be set in order to stop the iterations. Considering that it is computationally costly to know the number of gradient minima at each iteration, the practical criterion is that the great majority of pixels do not change significantly their value in the next iteration. 'Great majority' is set to 99.9% of pixels, whereas 'significantly' implies a difference of less than 0.05% of the dynamic range of the image. This means that for e.g. an 8-bit image, the threshold is set to 0.13 DN (note that the image is converted to floating point format as to allow decimal fractions). Finally, if the so defined convergence is not reached after twenty iterations (which typically occurs for low values of p), the process is stopped to prevent further simplification.

Although filters can be evaluated regarding their mathematical well-posedness (see e.g. the analysis of Weickert (1997b) on the Perona-Malik filter), the interest here is focused on the whether the filter does improve the baseline partition. The later has been derived from two sample images (1a, grey-level and 1b, colour composite, see 3.10) without applying any filter; and applying GIWEPS using four different values of p (Table 3-2). For p equal to 1 (and even 2 for the grey-level image), the resulting watershed partition is undersegmented, meaning that the simplification (reduction of gradient minima) effected to the image is excessive. Consequently the output granules are too big, and so is the RMSE. With lower diffusivities (greater p), the mean size of granules is in the same narrow range than the one obtained without filtering, and the resulting error is reduced.

filter	TM band 3			TM RGB 432		
	Watershed Partition	**Baseline Partition** MMU=5 Ha (80 pixels)		**Watershed Partition**	**Baseline Partition** MMU=5 Ha (80 pixels)	
	mean size	**mean size**	**nRMSE**	**mean size**	**mean size**	**nRMSE**
none	9.03	225.95	0.3728	5.62	221.84	0.2760
GIWEPS p=1	97.89	530.75	0.5072	29.47	190.51	0.2774
GIWEPS p=2	61.59	333.04	0.4101	6.57	209.82	**0.2679**
GIWEPS p=4	14.85	239.93	**0.3679**	5.91	220.711	0.2725
GIWEPS p=8	10.64	216.39	0.3692	5.81	230.75	0.2745

Table 3-2. The impact of different filtering choices in the error of the baseline partition.

The best results are obtained for p=2 (colour composite) and p=4 (grey-level). The diffusivity required for the colour image is greater because the surrogate gradient magnitude image is more intricate than the one of the grey-level image, since it accounts also for chrominance variations. Although a more thorough analysis is required for an optimal tune-up of the filtering stage, these values will be used tentatively in the current implementation.

3.7. Watershed partition

Following the analogy of the previous chapter, the area of influence of each gradient minimum forms a basin of attraction that is bounded by singular pixels. Singular pixels are those that are influenced by more than one attractor, and thus it cannot be reliably established to which basin they belong. The set of singular pixels defines the morphology *K* of the image, i.e. a complete partition that can also be seen as the *primal sketch* sensu Marr (1982b) of the original image. Therefore, the problem of obtaining the first partition in the baseline method can be reduced to locate such singular pixels.

In order to perform this task, a new analogy is introduced that has led to a powerful segmentation method: the watershed transform. Its origins can be dated back to an early contribution of James Clerk Maxwell (1870) to Geodesy:

" .. each point of the earth's surface has a line of slope, which begins at a certain summit and ends in a certain bottom. Districts whose lines of slope run to the same bottom are called Basins or Dales... Hence the whole

earth may be naturally divided into Basins or Dales, each point of the surface belonging to a certain dale... Dales are divided from each other by Watersheds..." (Maxwell 1870).

The application of these topographic concepts to the field of image analysis was introduced by Beucher and Lantejoul (Beucher & Lantuejoul 1979). Two decades later, Vincent and Soille (Vincent & Soille 1991) proposed an efficient algorithm for the implementation of the watershed transform, which is the one that will be adopted here. The idea of such transform is to consider the gradient magnitude image as a Digital Elevation Model (DEM), i.e. a square grid superimposed on a virtual territory where the value at each cell is the mean altitude of the terrain corresponding to that cell. The goal is to find the drainage divides, or watersheds, of that territory. The watersheds define a network of ridges that enclose the dales, or catchment basins, where each drop of rain would drain.

Note that the gradient magnitude image represents a DEM from a peculiar landscape, that could be assimilated to a lunar plain full of craters with ridges of different heights, where each crater corresponds to a homogeneous region in the original image, darker, brighter or of a different hue than its surroundings, i.e. a blob. Now it is clear the analogy with the concepts introduced in the previous chapter (table 3-3): stable attractors, associated to blobs in the image and to patches in the territory, are the gradient minima that remain when the filter reaches convergence (i.e. when the image graph reaches a steady state). Their basins of attraction are the catchment basins of those gradient minima. And finally, singular pixels defining the morphology of the filtered image are the watershed pixels of the output partition. An empirical evidence of the validity of this analogy is given at the end of this section.

Morphological domain	**Image domain**	**Geographic domain**
attractors	gradient minima	bottoms
basins of attraction	catchment basins	dales
singular points	watershed pixels	ridges

Table 3-3. The double analogy underlying the watershed transform

The watershed algorithm simulates a gradual immersion of the territory. Suppose that the bottoms (gradient minima) of craters are springs where pressurized underground water upwells. Then the water will begin to flood areas adjacent to the spring. Suppose further that the flow at each spring is such that the altitude of the water plane of the submersed areas is the same for all the territory (hence the analogy with immersion rather than flooding). Now, in places where the water coming from two different bottoms would merge, we build a dam of 1-pixel thickness, slightly taller that the highest crater of the territory. When the latter is completely submersed, we stop the flooding. The resulting dams are the watersheds of the territory, which in turn define a complete partition of the image.

Then the watershed algorithm can be summarized as follows:

1) Locate the set M of local minima of the gradient magnitude image G (see last paragraph of 3.5).
2) Sort the set M={1,...,i,...,k} of minima by magnitude in ascending order. Let v be a function returning the value of each pixel of G, and p a function returning the position of each pixel in G.
3) Initiate the catchment basin CB of each minimum with itself, i.e. $CB_i = \{p(M(i))\}$, the set of watershed pixels with the empty set $W=\{\varnothing\}$, and the same with the sets of candidate positions CP of each minimum $CP_i = \{\varnothing\}$.
4) Set the immersion altitude h to $h=\min(G)+\Delta h$, where Δh is a positive increment appropriately small.
5) Main loop: For i=1,...,k do the following:

   Check whether M(i) fulfils the next two conditions:

   5.1) $v(M(i)) \leq h$

   5.2.) $E(CB_i)=0$, where the E is a binary function that returns 1 if a CB is completely enclosed by watershed pixels, 0 otherwise.

   If the above conditions hold, include in CP_i positions p(j) such that:

   a) $v(j) \leq h$,

   b) $(p(j) \notin CB_h \;\; \forall h \neq i) \wedge (p(j) \notin W)$

   c) $p(j) \rightarrow p(M(i))$

   where condition c) means that the pixel j can be connected to the minimum under study through a path of pixels fulfilling a) and b).
6) Select those positions that have been included in more than one CP, if any. Put them in W and delete them from the respective CP sets.
7) Put the elements of all non-empty CPs in the corresponding CBs. Do $CP_i = \{\varnothing\} \; \forall i$.
8) Set $h=h+\Delta h$ (new iteration).
9) Repeat 5) to 8) until $h > \max(G)$.

Among the advantages of this algorithm with respect to other segmentation methods, it can be cited that 1) it captures global properties of the image, since watersheds cannot be identified locally (Olsen & Nielsen 1997); 2) watershed boundaries are always guaranteed to be connected and closed, whereas edge-based methods require complex erasing or connection of dangling edges (Gauch 1999); and 3) unlike most region-based methods, it needs no parameter from the user, like a similarity threshold or a stop condition. The main drawback is the profuse number of segments that it yields. Since the interest is usually focused in structures bigger than blobs, this is regarded as a disadvantage (oversegmentation). Therefore the watershed partition has to be simplified further.

There are three common ways of achieving such simplification: a) embedding the watersheds in a linear scale-space framework; b) thresholding the flood dynamics of watershed arcs (see below); and c) merging of catchment basins according to a similarity criterion (region merging) based on their mean signatures. The first two methods are discussed below, and the third is treated in the next section.

The first method (Jackway 1996;Olsen & Nielsen 1997;Gauch 1999) consists in creating a family of increasingly blurred images from the original one. This family forms a scale-space representation (Lindeberg 1994), in which the scale of each image is given by the width of the gaussian filter used to blur it. As a result of the blurring, successive images are simpler, meaning that there are less gradient minima. Then the watershed partition is computed at each scale, and the resulting catchment basins are linked across scales, yielding a hierarchical partition. This method is discarded for two reasons: first, it is computationally intensive, and second, the linkage is not as straightforward as one might think. Gaussian blurring deforms the structure of the original image, not only annihilating but displacing both edges and gradient minima. Therefore there may be situations where the assignment of previous regions to new ones cannot be clearly established. These matching ambiguities between successive images, that occur as well for pyramids (stack of images of increasingly larger pixel size), are known as the correspondence problem in computer vision (Cox 1993).

If the goal is to foreground the most contrasted objects within an image, the concept of geodesic saliency of watershed contours (Najman & Schmitt 1996) provides a powerful tool. To see it, imagine now that we simulate, instead of an immersion, a uniform downpour over the lunar territory represented by the gradient magnitude image. The first craters to be completely flooded would be the ones with the lowest ridges. More precisely, since ridges have not a uniform altitude, water will begin to overflow from saddle points (passes across the ridges). Therefore, for each ridge (watershed arc) separating two catchment basins, the higher the lowest of the saddle points in the arc, the later the overflow will take place. Then, as the downpour proceeds, basins separated by low passes will become connected and subsumed under the water. Using the altitude of these saddle points as an index to flood dynamics of watershed arcs, a hierarchical partition can be obtained.

However, this approach is not suitable for obtaining the baseline partition. The latter is characterized by having regions of uniform size, of the same order of magnitude than the MMU size. In contrast, successive partitions corresponding to increasing thresholds of watershed dynamics will show an increasing disparity in size, with many small high-contrast regions littered throughout larger regions (see figure 3-10). Besides, subtle radiometric differences separating two basins may later become relevant (i.e. constitute the boundary between two polygons) provided they survive the region-merging step. Hence it seems preferable not to erase prematurely weak watershed arcs. Then the only remaining method to reduce the number of segments of the watershed partition, is region merging, which is explained in the next section.

Before finishing this section, it is worth making clear the link between watershed partitions and Thom's (Thom 1975) theory of attractors. The validity of the analogy that equates Thom's singular points to watershed pixels, will be showed with the following example. GIWEPS has been applied during 200 iterations with p=4 to image 1a (see 3.10). Apart from the filtered image, an ancillary image was also produced in order to know the minimum weight (of nine) used in the computation of the value of each pixel in the last iteration. The gradient magnitude was computed from the filtered image, and the watershed algorithm was applied to it. If the analogy is valid, every watershed pixel should have a nearly zero value in the minimal weight image, since singular pixels are by definition those that do not interact with some of their neighbours because they are not similar to them. Conversely, pixels belonging to catchment basins (the regular points of Thom's theory) interact with all their neighbours and hence should have a higher value in the above-mentioned image.

The results approximate this situation: the average weight g_{cr} (equation 3.6.3) of watershed pixels is 0.0005, with a maximum observed value of 0.0864 and 0.0025 as the 95% percentile. In contrast, the mean value for pixels belonging to catchment basins is 0.3599, and more than one third is above 0.95, meaning that all their neighbours have roughly the same DN. However, many of them (most of those adjacent to watershed pixels) show a low weight value. The reason is that the width of watershed lines is one pixel, whereas the boundary between two regions is two-pixels wide. This can be seen in the table below, which show the process from the original image to the watershed partition for a 5x5 pixel sample:

59	60	62	63	65	60,48	60,50	60,53	60,56	64,15
59	62	64	65	65	60,46	60,48	60,52	64,14	64,16
60	63	66	66	65	60,41	60,43	64,13	64,15	64,16
60	61	64	66	67	60,35	60,35	64,13	64,15	64,16
58	60	63	66	66	60,26	60,24	64,13	64,15	64,15
0,9975	0,9906	0,0080	0,0098	0,0102	11,09	11,02	0,10	0,16	0,16
0,9906	0,0054	0,0066	0,0068	0,0098	11,03	0,06	0,11	0,11	0,12
0,9789	0,0040	0,0029	0,0066	1,0000	10,91	0,05	0,05	0,08	11,11
0,9847	0,0028	0,0023	1,0000	1,0000	10,97	0,04	0,03	11,11	11,11
0,9853	0,0003	0,0016	1,0000	1,0000	10,94	0,01	0,02	11,11	11,11
0,0181	0,0218	0,0254	1,9946	1,9842	1	1	1	0	2
0,0314	0,0355	2,0120	2,0029	0,0109	1	1	0	0	2
0,0444	1,4584	2,0324	0,0131	0,0067	1	1	0	2	2
0,0587	1,4866	1,4891	0,0096	0,0055	1	1	0	2	2
0,0921	1,5263	1,5309	0,0090	0,0070	1	1	0	2	2

Table 3-4. A 5x5 pixel sample showing the process from the original image to the watershed partition.

The top left box comes from the original image, while the top right corresponds to the filtered one. The sample displayed consists of two regions of almost constant value, whose thin boundary is represented by a zigzagging line. The values of the minimum observed weight g_{cr} and w_{cr} (the latter expressed as a % of the summation of the nine weights allocated to each pixel) in the last iteration is shown respectively in the middle left and right boxes. Note that pixels having low minimal weights are those touching the zigzagging line separating the regions, i.e. those having at least one of their eight neighbours belonging to a different region. Finally, the watershed partition (bottom right box) is formed by locating the ridges (cells in grey) of the gradient magnitude (bottom left box) computed from the filtered image.

By applying a non-linear diffusion process to the image, it evolves to a piecewise constant image. This evolution can be viewed as grouping pixels into various perceptual units (blobs), corresponding to different basins of attraction (see the quotation below). What the watershed algorithm does is to draw the divides separating these basins. The resulting closed network of boundaries is the primal sketch of the filtered image, which in turn is a simplified version of the original one. If the goal is to reconstruct the objects portrayed in it, one has to identify first its minimal components, which consist of perceptual blobs that correspond to catchment basins.

It is important to stress that one important contribution of this thesis is the implicit conclusion that gradient watersheds are more than just one-among-many segmentation methods: they provide an operational tool with which to apply Thom's (1988) 'semiophysics' to image analysis[1]. If in the following excerpt, the word 'dissipation' is replaced by the 'diffusion' applied to the image prior to the watershed transform, a similar conclusion seems to have been drawn much earlier by Stephen Wolfram (1986) in a far-sighted article on the possible applications of the (to-be-established) principles of complexity to engineering problems:

"... Dissipation, in one of many forms, is a key principle which lies behind much of the robustness seen in natural systems... Such behaviour is typically represented by a differential equation whose solution tends to a fixed point at large times, independent of its initial conditions... This is the case for an idealized ball rolling on a landscape, with dissipation in the form of friction. Starting at any initial point, the ball is ``attracted'' towards one of the local height minima in the landscape, and eventually comes to rest there. The set of initial positions from which the ball goes to a particular such fixed point can be considered as the ``basin of attraction'' for that fixed point. Each basin of attraction is bounded by a ``watershed'' which typically lies along a ridge in the landscape. Dissipation destroys information on details of initial conditions, but preserves the knowledge of which basin of attraction they were in. The evolution of the system can be viewed as dividing its inputs into various ``categories'', corresponding to different basins of attraction. This operation is the essence of many forms of pattern recognition..." (Wolfram 1986)

Finally, note that the watershed partition can also be viewed as a rough zonation (2.2.14) of the imaged territory where the watershed pixels are the cells of the reference grid overlapping with more than one ground object (patch). The analogy is valid as long as the coincidence hypothesis (equating catchment basins –blobs- to patches) holds. Then watershed pixels are also equivalent to an epsilon band of fixed width where the 'true' boundary between patches lie.

[1] Note however that the topological catastrophe theory is by no means new in image analysis. In particular, it has been applied in the context of scale-space to study the evolution of critical points (either saddle points or local intensity extrema) across the family of increasingly blurred images. See (Kuijper & Florack 2001) for a review.

3.8 Region merging

The final step in the construction of the baseline partition is to aggregate the regions of the watershed partition (blobs) into regions exceeding the MMU size (granules). There are many ways in which this task can be performed, that can be generically categorized as region merging methods, where the seeds are catchment basins. Region merging algorithms can be defined according to the following features:

1) the way the initial regions are chosen.
2) the similarity measure(s) used to merge regions.
3) the merging procedure (threshold and merging order).
4) the stop criterion.

The method presented here does not pretend to be the best possible choice, it is just designed to demonstrate the type of results that can be achieved by applying the concepts set forth in the previous chapter. Therefore it will not be compared to other methods. For a recent review on image segmentation, see (Cufi et al. 2002). For classical ones, see (Haralick & Saphiro 1985) or (Pal & Pal 1993). Notwithstanding the above statement, it is worth noting that a segmentation sequence consisting of image smoothing and/or gradient magnitude simplification, watershed transform plus region merging, has already been used in different contexts (Haris, Efstratiadis, & Katsaggelos 1998;Weickert 1998;Fjørtoft et al. 1998;Ji & Park 1998). However, none of these studies was related to landcover mapping. Besides, except for the watershed transform, the methods proposed here are new and not based upon the ones used by those authors.

The most distinctive feature of the algorithm proposed in this thesis, hereafter called *size constrained region merging* (SCRM), is that it imposes no threshold on the similarity of the regions to be merged. Rather, the merging is based on whether the involved regions are greater than the specified MMU. Actually this is the stop criterion in SCRM: the merging proceeds until all the regions in the partition are larger than MMU. Many region merging algorithms include a size constraint (see e.g. (Hagner 1990;Woodcock & Harvard V.J 1992;Baraldi & Parmiggiani 1994), but of all them set thresholds on the dissimilarity measure. In turn, SCRM allows disparate regions to merge, but the merging sequence is programmed in such way that the homogeneity of the resulting regions is maximal given the size constraint.

The idea of SCRM is quite simple: the most similar neighbour to each segment is identified, and then, beginning by the segments that show the highest similarity, the segments are merged iteratively until all the segments are larger than the MMU, allowing only one merge per segment and iteration, and not allowing aggregation when i) both segments are already larger than MMU, ii) some neighbour of one of the segments has already been merged in this iteration, and iii) one of both is smaller than MMU but its most similar neighbour is radiometrically closer to it than to the segment under evaluation. By permitting only one merge per pass, it is implicitly assumed that similarity is not transitive (i.e. if B is similar to A and C it doest not necessarily imply that A and C are similar, they could be e.g. in opposite sides of B in the data space). Constraint ii) is imposed because every time a segment is merged, the resulting merger may potentially become the most similar segment to some of its neighbours, and hence similarity has to be reassessed. Finally, by enforcing constraint iii), the resulting segments are expected to reach the highest possible homogeneity given the size constraint. That is to say that homogeneous regions are formed first, and then dissimilar gaps smaller than MMU are progressively incorporated to the former.

The SCRM algorithm can be stated as follows:

1) Get the list lb of labels of the n segments of the current partition (the watershed in the case of the first iteration), such that lb(i) returns the numeric label of the segment in position i of the list. After the first iteration, the list is given by suppressing repeated labels in nwl (see below). Maintain a link between the initial segments and new ones, so that at the end the final label fl(i) of each initial segment is known.

2) Get the size of each segment i in lb, and store them in an n-elements array sz such that sz(lb(i)) returns the number of pixels inside segment lb(i). After the first iteration, sz(lb(i)) is given by summing up the size of the segments of the previous partition that constitute a new segment.

3) Get the mean signature of each segment lb(i), and store it in an n×m array of signatures, such that the function sg(lb(i)) returns the m-dimensional signature of segment lb(i). In the first partition, segment signatures are obtained from pixels inside them, and in successive ones, from the weighted (by size) mean of the original segments within it.

4) Compute the adjacency list adj (adj is a one-dimensional array (vector) that is ordered in such a way that for each segment lb(i) of n, the list of neighbours of lb(i) is given by N(lb(i))=adj[adj[lb(i)]:adj[lb(i)+1]-1], where adj[*l*:*m*] is the subset of elements of adj included between the positions *l* and *m*, both inclusive).

5) Create three additional vectors of n elements: msn, to store the label of the most similar neighbour to each segment, md, to store the corresponding normalised distance, and nwl, to store the new labels of segments after the iteration.

6) For each segment lb(i) from i=1,...,n. do the following

   Compute $NVD(sg(lb(i)),sg(N(lb(i))_j))$, with j=1,..k, where $N(lb(i))_j$ is the label of the j^{th} neighbour of segment lb(i), and NVD is the normalised vector distance (3.4).

   Select the neighbour g such that

   $$NVD(sg(lb(i)),sg(N(i)_g))<NVD(sg(lb(i)),sg(N(lb(i))_j)) \quad \forall j\neq g.$$

   Do $msn(i)=N(lb(i))_g$ and $md(i)=NVD(sg(lb(i)),sg(N(lb(i))_j))$.

7) Sort md in ascending order. Sort msn and lb using as an index the previous sorting, as to keep the correspondence between those vectors.

8) Create a binary checklist chk, with chk(lb(i))=1 meaning that the segment lb(i) is allowed to merge in this iteration, and chk(lb(i))=0 otherwise. Do chk(lb(i))=1 ∀i.

9) For each segment lb(i) from i=1,...,n

   if chk(lb(i))=1 and sz(lb(i))<MMU and chk(msn(i))=1 then do the following:

   nwl(i)=msn(i)

   For j=1,...,k , with k equal to the number of neighbours of lb(i), do the following (enforcement of constraint ii):

   $chk(N(lb(i))_j)=0$

   For h=1,...,k, with k equal to the number of neighbours of $N(lb(i))_j$, do the following (enforcement of constraint iii):

   If $md(N(lb(i))_j) < md(N(N(lb(i))_j)_h)$ then $chk(N(N(lb(i))_j)_h)=0$

For j=1,...,k, with k equal to the number of neighbours of msn(i), do the following (enforcement of constraint ii):

chk($N(msn(i))_j$)=0

For h=1,...,k, with k equal to the number of neighbours of $N(msn(i))_j$, do the following (enforcement of constraint iii):

If md($N(msn(i))_j$) < md($N(N(msn(i))_j)_h$) then chk($N(N(msn(i))_j)_h$)=0

10) Update the mapping fl using nwl.

11) Repeat 1) to 9) until there are no mergers after current iteration.

12) For i=1,...n_0 do px(l_0(i))=fl(i), where n_0, l_0, and fl are respectively the number of segments, the labels and the final labels of the segments of the initial (watershed) partition, and px is the set of pixels that belonged to segment l_0(i) of the initial partition.

13) Fill watershed (0-valued) pixels of arcs lying in the interior of final segments with the numeric label of the corresponding segment.

Due to the fact that the signature, size and adjacency of segments are derived from previous lists, the image can be scanned only once, and the labelled image representing the partition only twice (at the beginning and at the end). Therefore the algorithm is quite fast under this implementation, taking only a few seconds to complete. However, the output may be suboptimal, since there are a number of interior pixels that are left out of signature calculation because they were watershed pixels in the first partition. Besides, size is underestimated for the same reason. An amendment is introduced that consists in erasing superfluous watershed arcs (i.e. performing steps 12 and 13) each time the number of segments in the partition is halved, i.e. when most of the segments have been merged at least once since last time the partition was rebuilt. In this event, the size and signature of the remaining segments are recomputed in the same way as in the first partition. With this modification, the mean size of granules is slightly reduced and so is the resulting nRMSE. However, this procedure has an unwanted side effect on the reproducibility of the results (see 3.11).

3.9 Vectorization

Once the baseline partition is obtained, the last step is to convert it into a vector layer. Through vectorization, the size of the representation of the baseline partition is drastically reduced in comparison to the raster format. Hence raster to vector conversion follows the Minimum Description Length (MDL) Principle (Rissanen 1978b). In order to proceed, the centres of the remaining watershed (0-valued) pixels are considered the initial vertices conforming the vector layer. Note that this is analogous to consider watershed pixels as a transition zone between patches that can be represented by its medial axis. The nodes (junctions) are identified, so that each vector unit (polygon) is defined by the set of watershed arcs bounding the corresponding granule. Finally, superfluous vertices (those roughly lying on the same line connecting their preceding and succeeding neighbours) are deleted. Once the partition is vectorized, an associate database is filled with radiometric information (mean in each band) about each granule. The procedure is as follows:

1) Initialize the vector layer and the associate database. For each granule i from i=1 to n do the following:
2) Get the set of 0-valued pixels bounding the granule. Order the set in topological succession, so that the list begins and ends with the topmost junction pixel.
3) Transform the previous set in a two-column array in which each row consists of the cartographic coordinates of the centre of each pixel in the list.
4) For j=1,...,k delete from the list those vertices v_j such that $d(v_j, l(v_{j-1},v_{j+1}))<psz$, where $l(v_{j-1},v_{j+1})$ is the straight line connecting the preceding and succeeding neighbours of v_j, $d(v_j, l)$ is the orthogonal distance between that line and v_j, and psz is the pixel size.
5) Add the output of 4) as a new record to the vector layer.
6) Compute the mean of the granule en each band, as well as its area, perimeter, and geographic coordinates of its centre. Put the result in the corresponding record of the associate database.

In the example, the raster baseline partitions are in the order of MB, while the corresponding vector layers are in the order of kB.

3.10. Examples

In this section, some results obtained with the baseline method are shown. It has been applied to two sites. Image 1a and 1b come from a Landsat 5 TM subscene of 10x10 km^2 (centre coordinates 40º 54' N, 2º 18' W), acquired on 06 July 1994 over a territory in the Guadalajara province, Spain. Image 1a correspond to TM band 3 (red), whereas image 1b is a RGB 432 false colour composite. The image was ortho-rectified to UTM projection using the geo-correction tool from Erdas™ and a Digital Elevation Model (DEM), and resampled to a 25m pixel size (same of the DEM) through cubic convolution. No radiometric correction was performed. This site is covered by deciduous (red) and evergreen (brownish red) oak shrubland, pine forests (dark red areas in top and right part of the image), juniper sparse woodland (dark blue), thickets (greyish blue) and agricultural fields (light green and blue, beige, and bright red).

Image 1a is shown with no overlay (**Figure 3-3**), and with the vector layer of the baseline partition with MMU=5 ha, obtained without any filtering (**Figure 3-5**), and applying GIWEPS with p=4 (**Figure 3-6**). The latter is also displayed (**Figure 3-7**) with the output of applying the same parameters to a 200x200 pixel subset, which is zoomed in **Figure 3-8**.

Image 1b is shown with no overlay (**Figure 3-4**), with the original MFE vector layer (**Figure 3-17**), and with MFE polygons delineated with the outer edges of the segments (of the baseline partition of figure 3-15) encompassed within each one of them (**Figure 3-16**). The watershed partition of the gradient magnitude of image 1b is displayed in **Figure 3-9**. Additionally, the result of suppressing watershed arcs of low dynamics (Najman & Schmitt 1996) is shown in **Figure 3-10**. Successive figures display the baseline partition obtained for several values of the MMU: 2 ha (**Figure 3-11**), 5 ha (**Figure 3-12**) and 10 ha (**Figure 3-13**). Finally, the 10 ha baseline partition derived using i) the six non-thermal bands of the Thematic Mapper **(Figure 3-14**) for this subscene and date; and ii) the first three principal components of it (**Figure 3-15**), are also shown.

Figure 3-18 displays the result of an unsupervised classification carried out on the mean signature from granules of the baseline partition from figure 3-15 (10 ha, PC 123). The data space (three bands) was divided into fifteen spectral classes with the K-means clustering algorithm (Hartigan & Wong 1979). Using the information of the MFE, spectral classes were later aggregated meaningfully into four landcover types: agriculture (displayed in a mint cream colour), thickets (light blue), sparse woodland (slate grey) and forest (sienna). A more detailed account was not possible since each spectral class included semantically different zones at a lower level of abstraction. Each granule is displayed with the colour of the landcover type allocated to its mean signature. MFE polygons are also overlaid for comparison.

Image 1c is a Landsat 5 TM 432 RGB composite of the same subscene acquired in December (Dec 6 1994) with a sun elevation angle of 21º. The baseline partition was computed for a MMU of 10 ha and superimposed to the image in **Figure 3-19**. Image 1d (**Figure 3-20**) is a shaded relief computed from a DEM corresponding to the subscene, and using the same solar elevation and azimuth than image 1c. The previous partition has been overlaid on it as to allow comparison.

Finally, image 2 is a 4x4 km^2 panchromatic aerial ortho-photo (UTM) of 1m-pixel size acquired in October 1997 over of a site (centre coordinates 39º 54' N, 0º 22' W) in the Valencia region, Spain, vegetated with cork trees (darker region) and *garrigue* (Mediterranean shrubland). The baseline partition (MMU=2 ha) was derived from the original image (**Figure 3-21**) and from a coarser version resized by pixel averaging to 10 m (**Figure 3-22**).

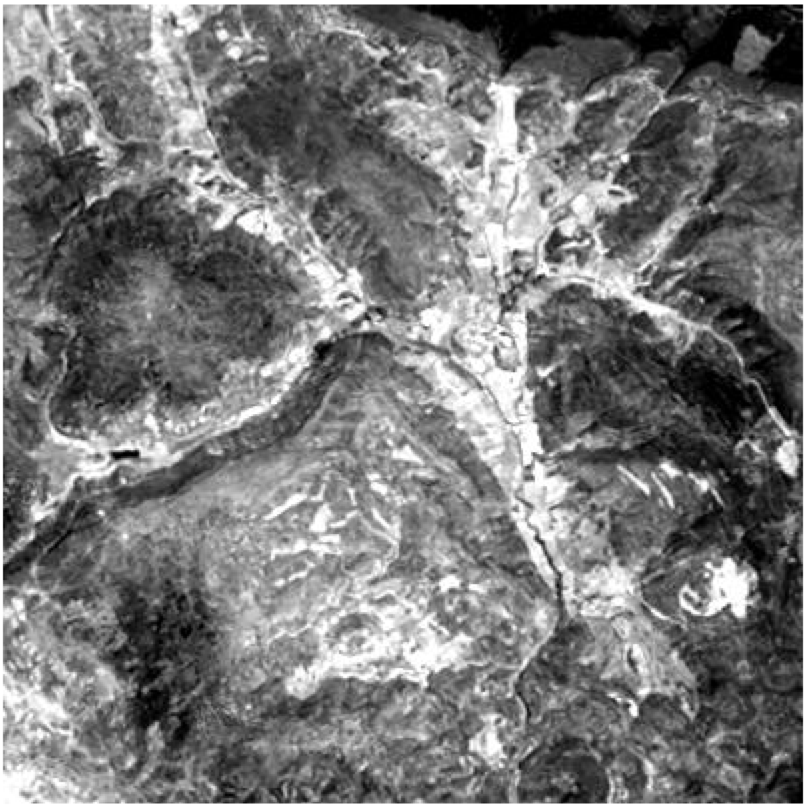


Figure 3-3. Image 1-a (TM band 3)

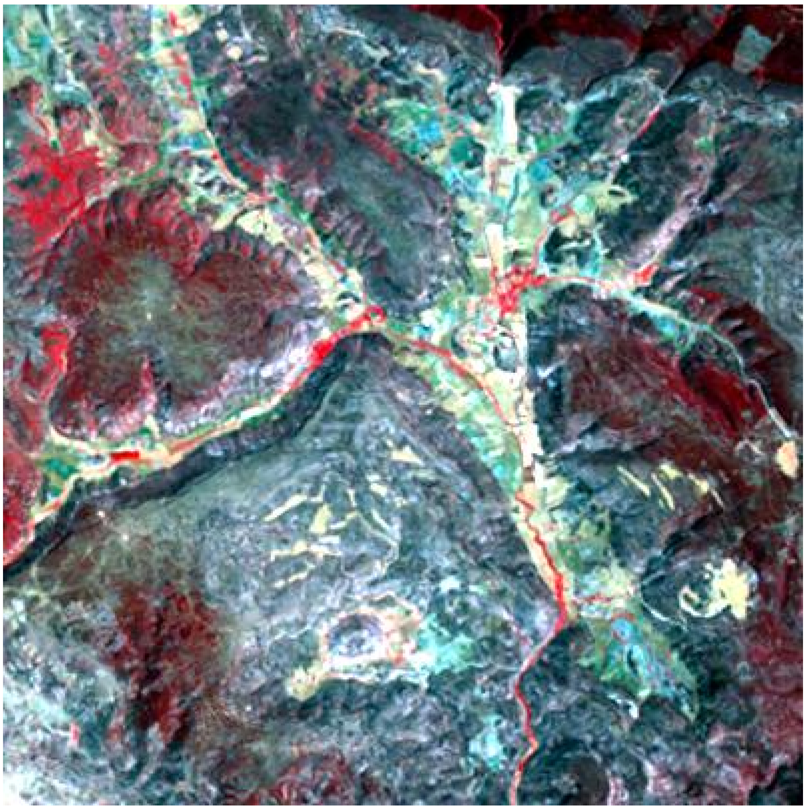


Figure 3-4. Image 1-b (TM RGB 432)

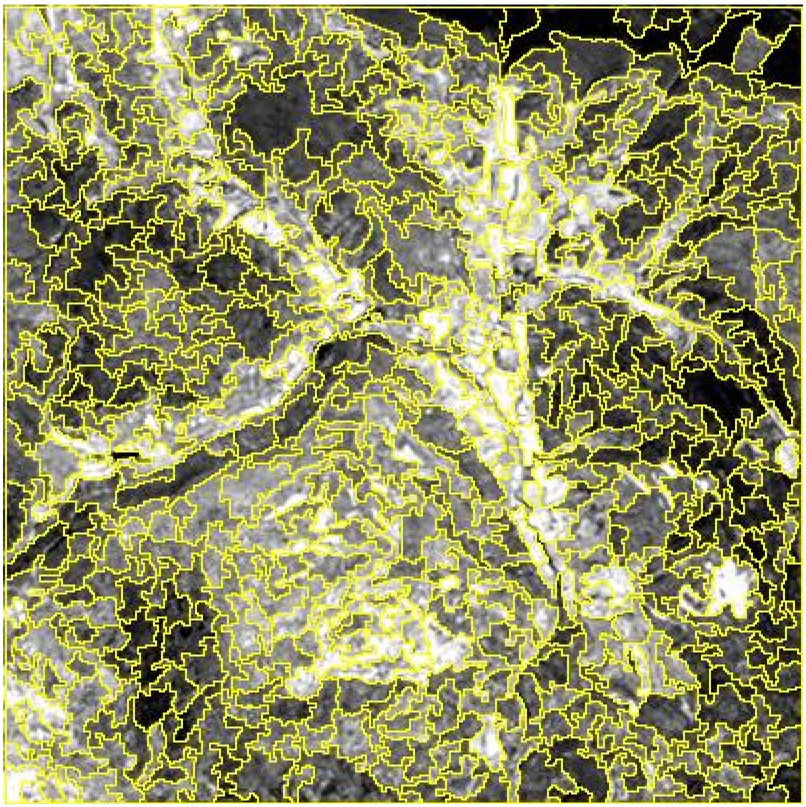


Figure 3-5. Image 1a: Baseline partition (MMU=5 ha) with no filtering

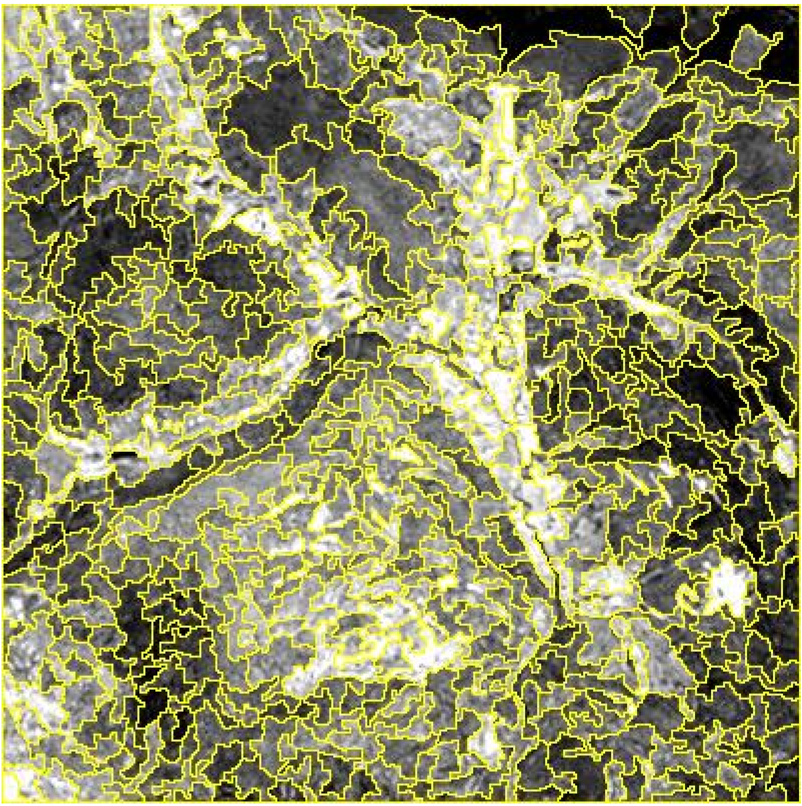


Figure 3-6. Image 1a: Baseline partition (MMU=5 ha) with GIWEPS (p=4)

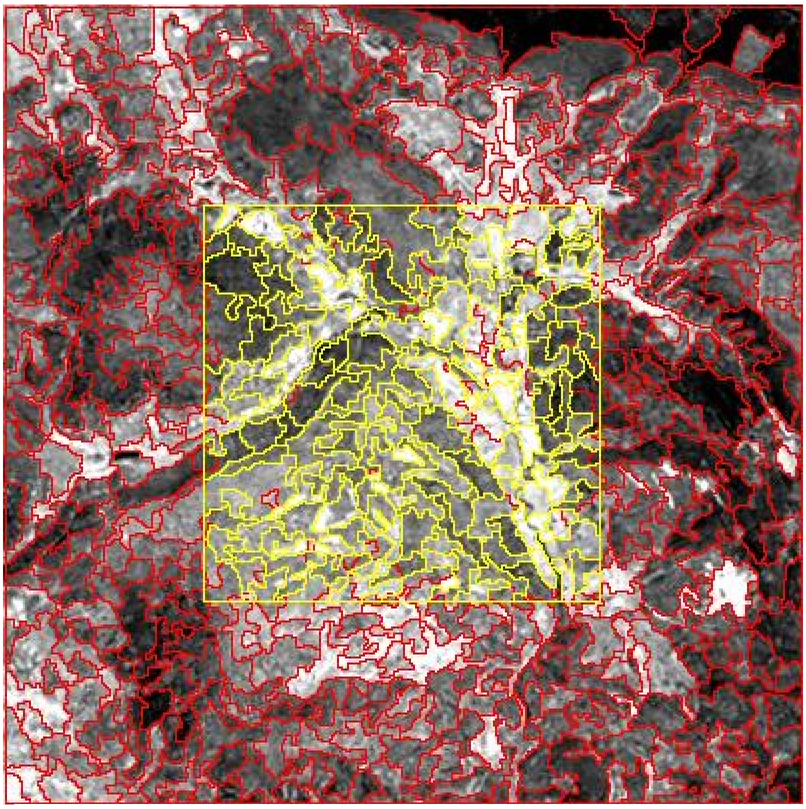


Figure 3-7. Image 1a. Red: previous partition (figure 3-6) ; Yelow: output of computing the partition (same MMU and p) in a 200x200 pixel subset

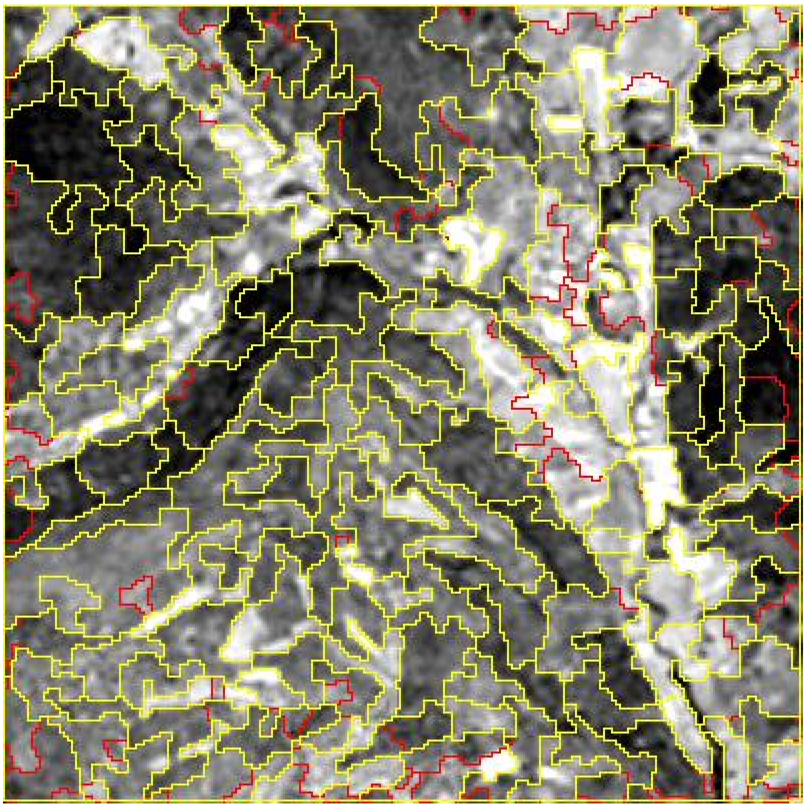


Figure 3-8. Zoom of figure 3-7 showing arcs altered (red) due to the discontinuous physical update of the partition during merging

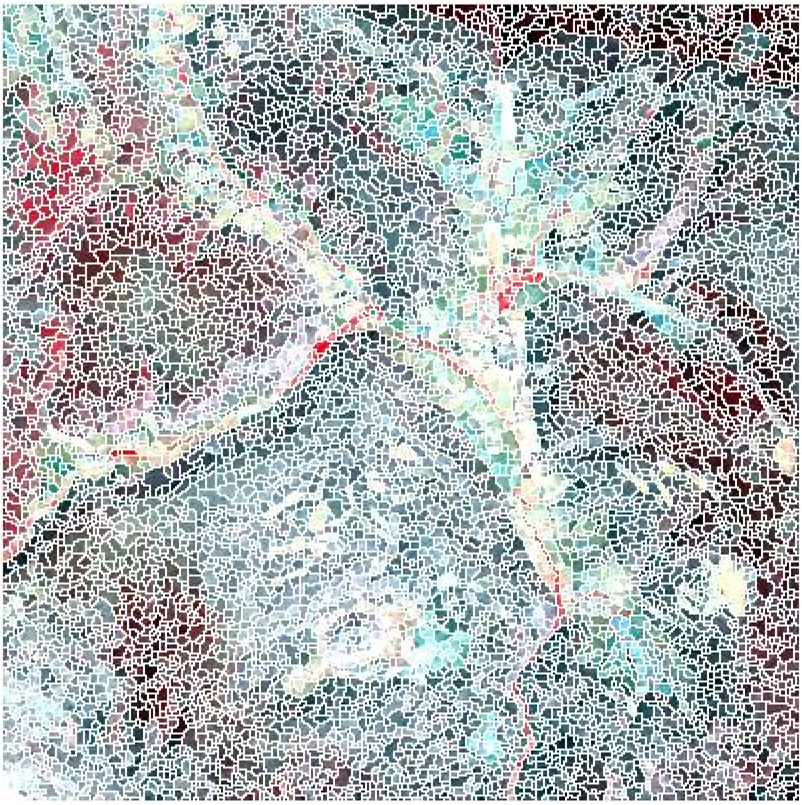


Figure 3-9. Image 1b: watershed partition

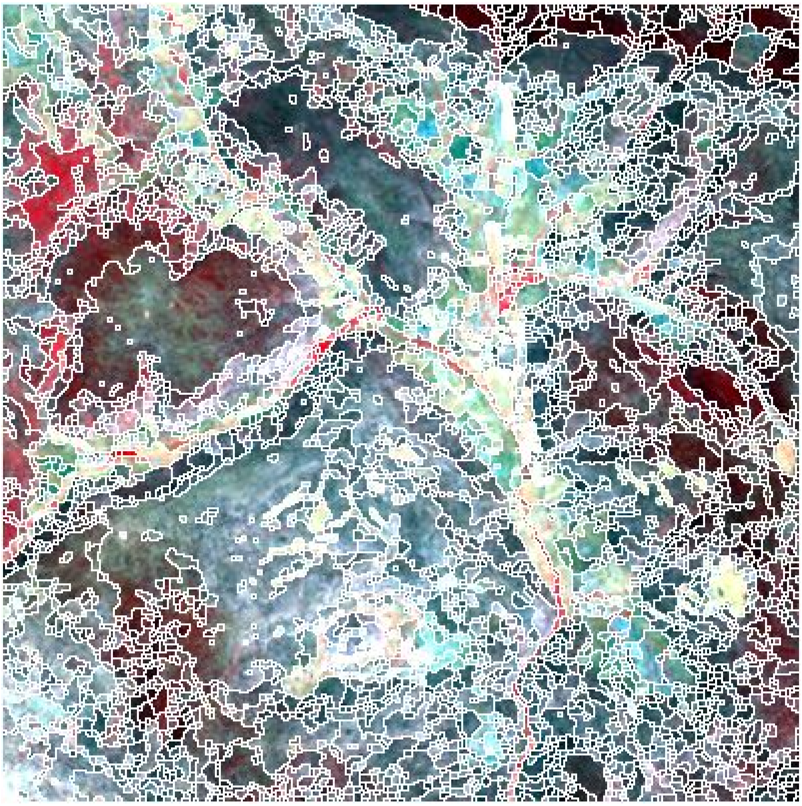


Figure 3-10. Image 1b : watershed partition after suppressing arcs of low dynamics

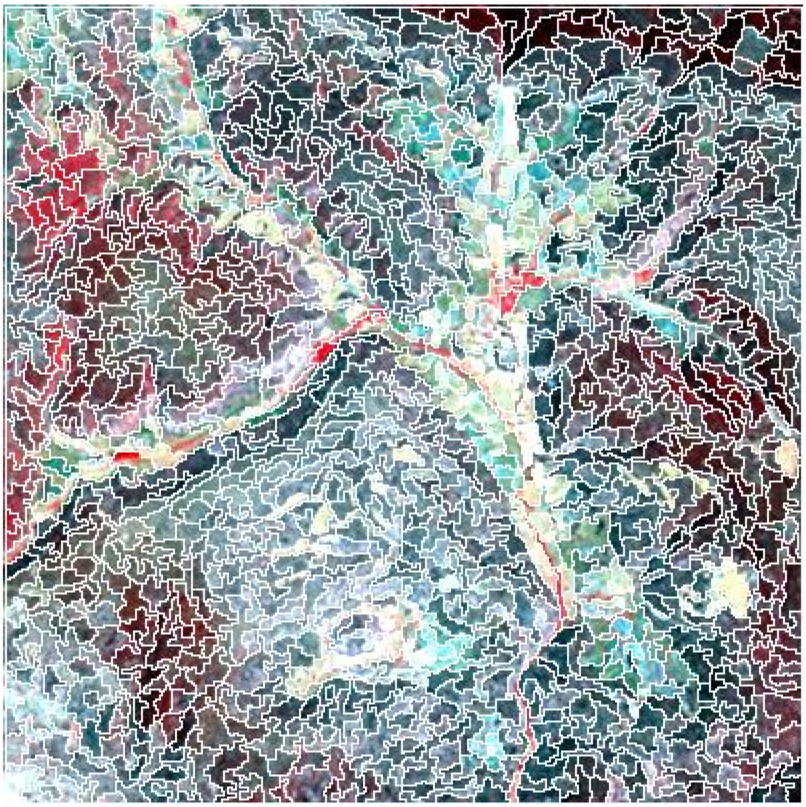


Figure 3-11. Image 1b: baseline partition (MMU= 2 ha)

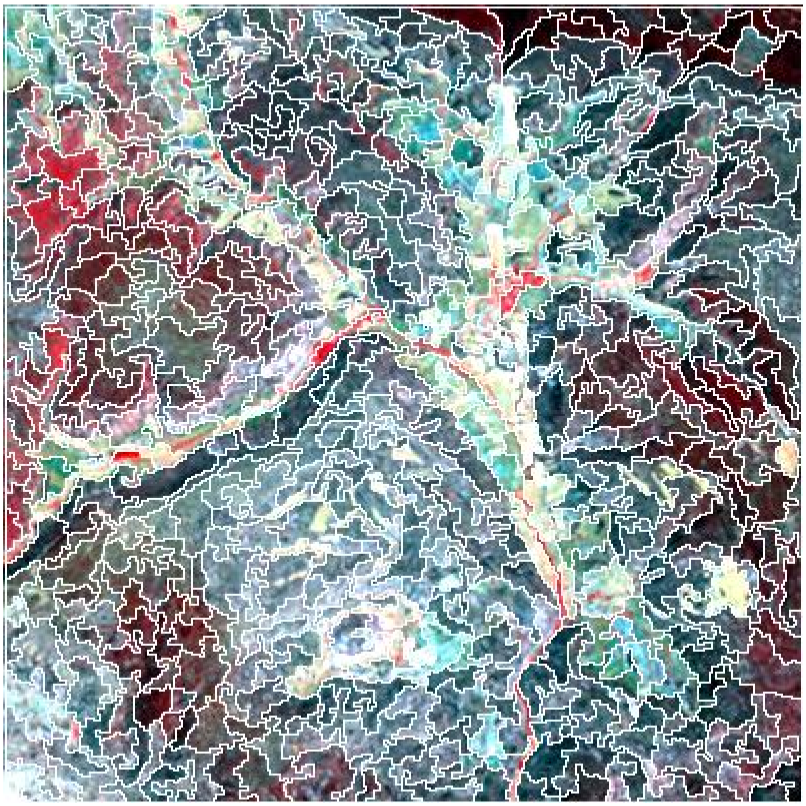


Figure 3-12. Image 1b: baseline partition (MMU=5 ha)

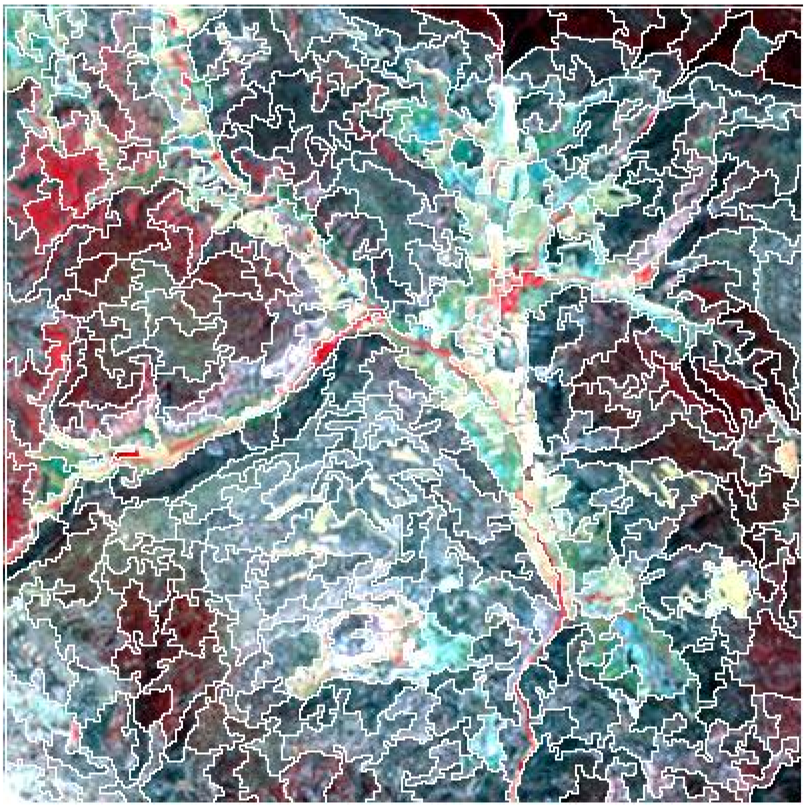


Figure 3-13. Image 1b: baseline partition (MMU = 10 ha)

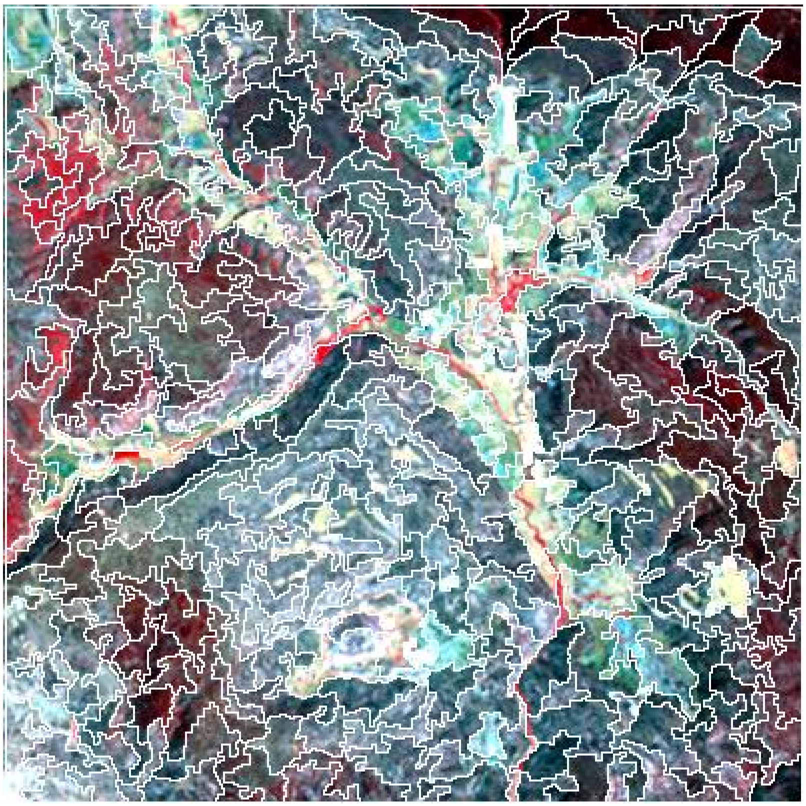


Figure 3-14. Image 1b: baseline partition (10 ha) using 6 TM bands

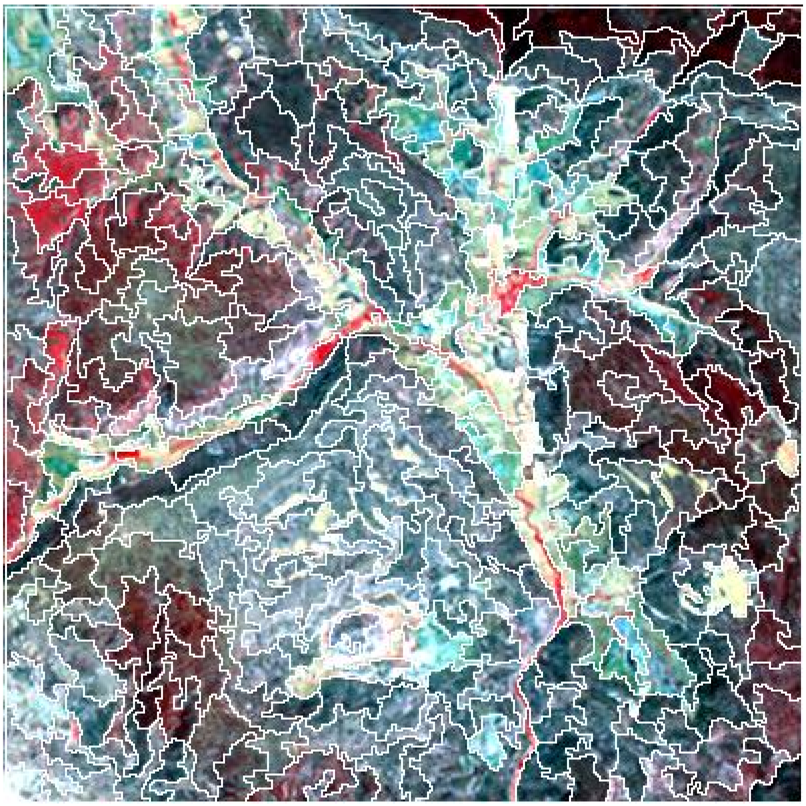


Figure 3-15. Image 1b: baseline partition (10 Ha) using TM Principal Components 123

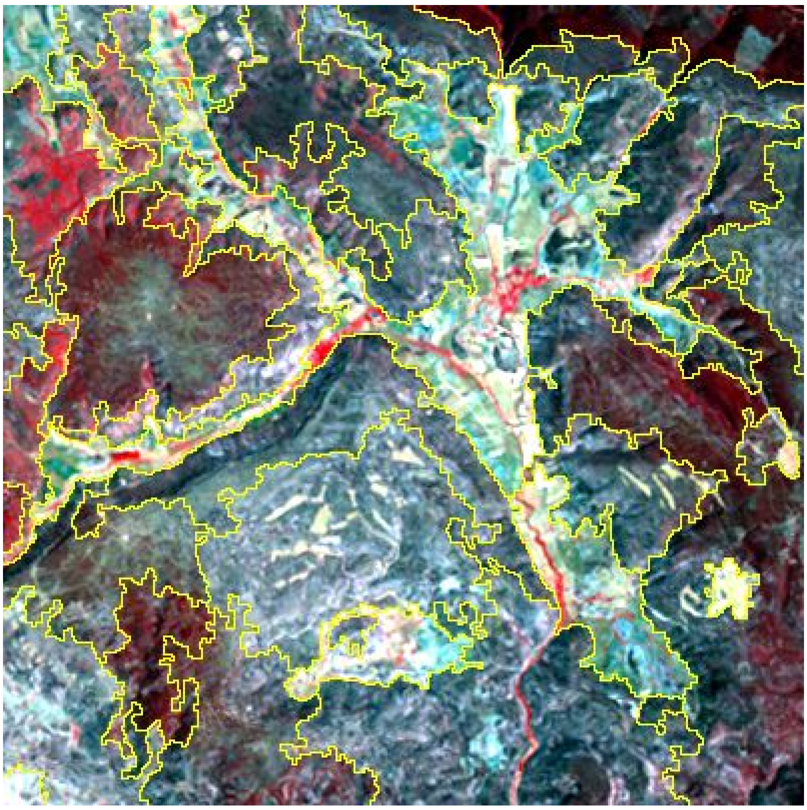


Figure 3-16. Image 1b: Reconstruction of the MFE polygons according to maximum overlap with the baseline partition of the previous figure

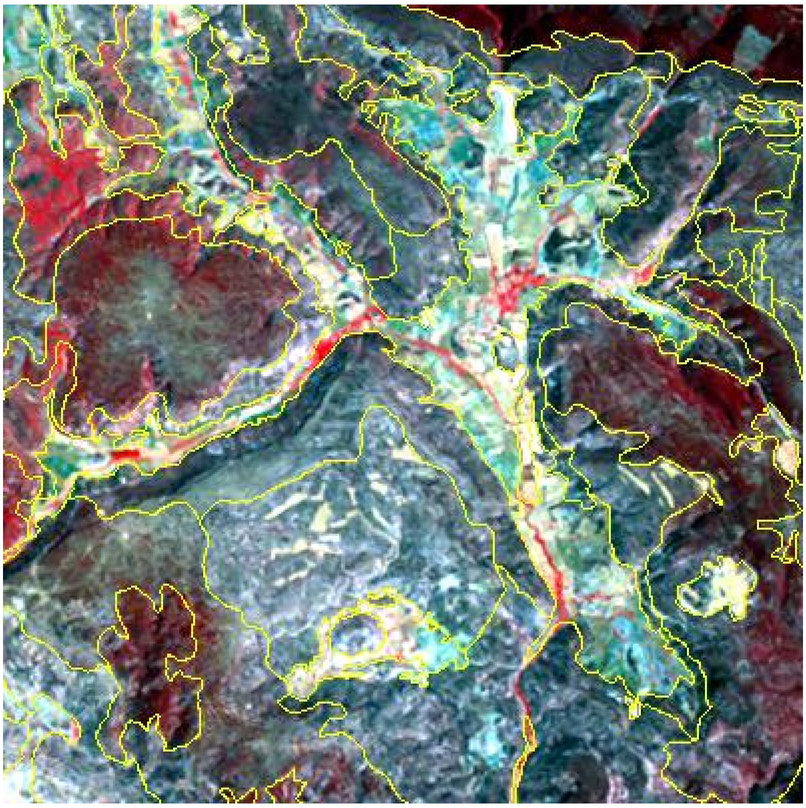


Figure 3-17. Image 1b : original MFE polygons

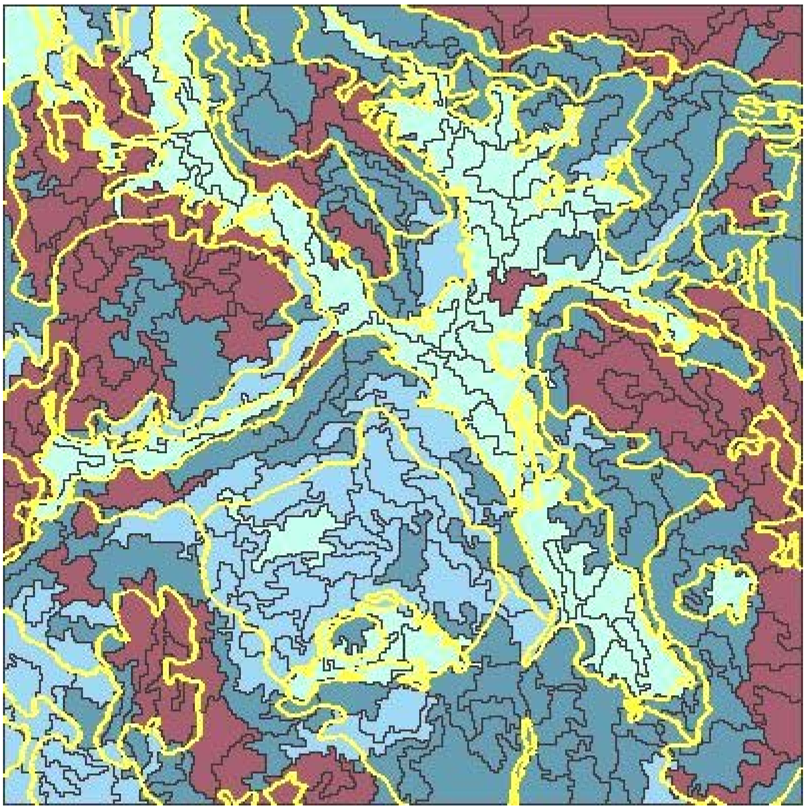


Figure 3-18. Broad unsupervised classification of the segments of figure 3-15 (baseline partition 10 Ha using TM PC123) with the MFE polygons overlaid

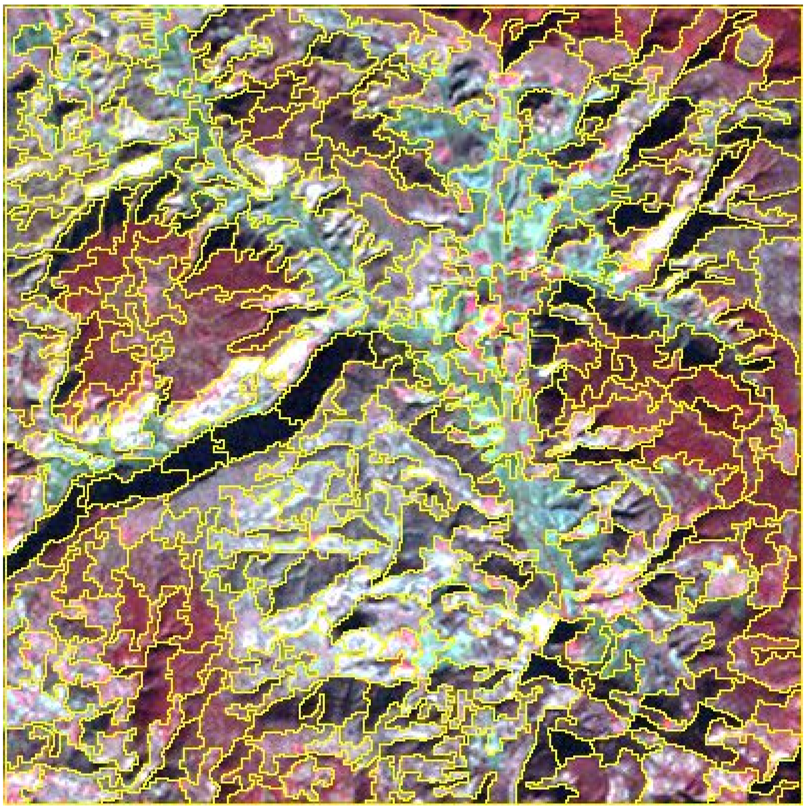


Figure 3-19. Image 1c (TM 432 RGB winter): baseline partition (10 Ha)

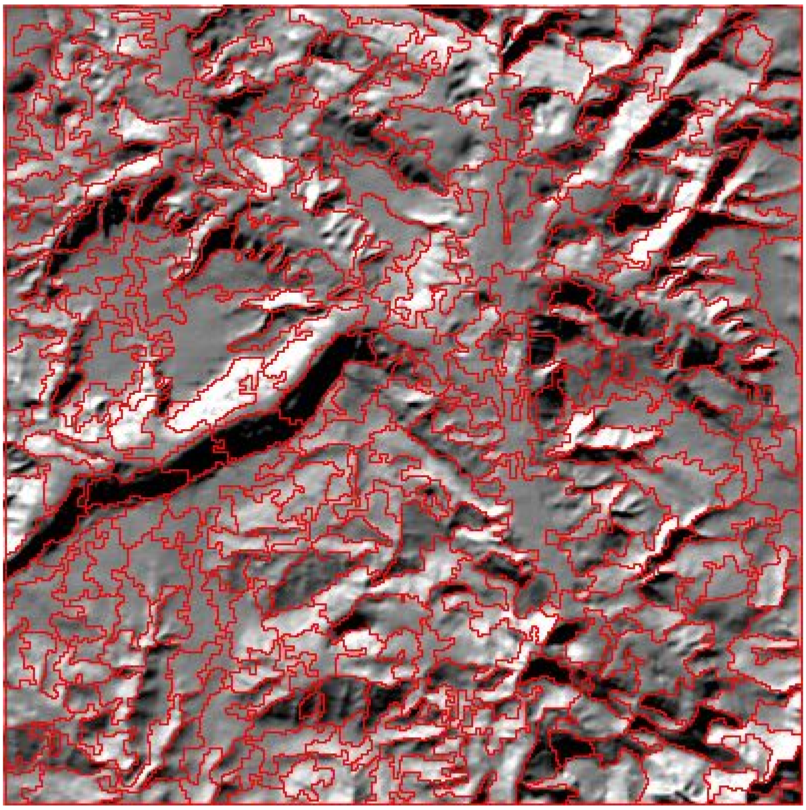


Figure 3-20. Image 1d: DEM shaded relief with the same solar elevation and azimuth than image 1c, and with the same overlay .

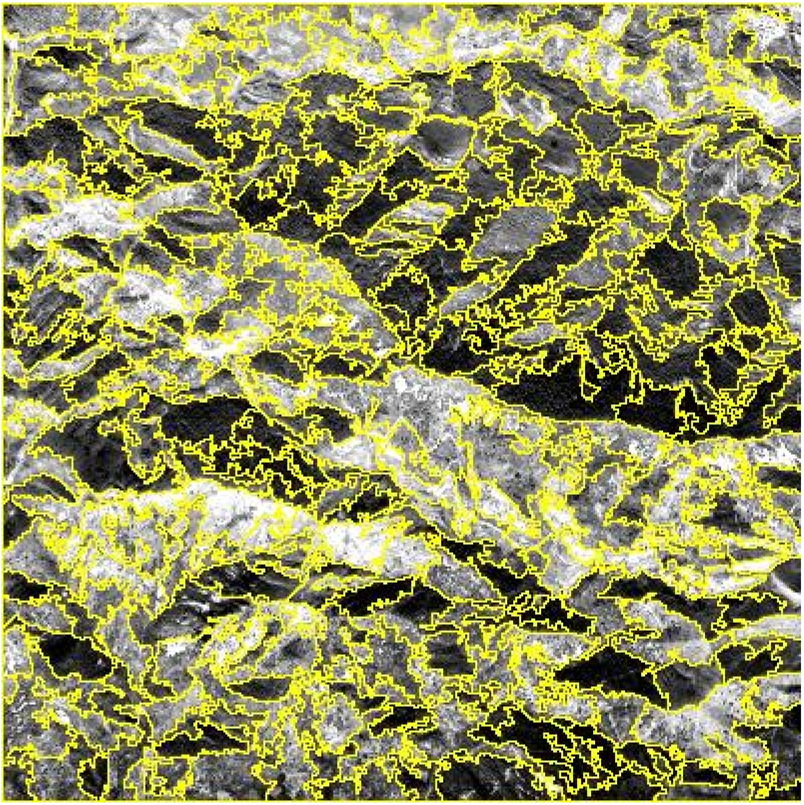


Figure 3-21. Image 2: baseline partition (MMU=2 ha, pixel size=1m)

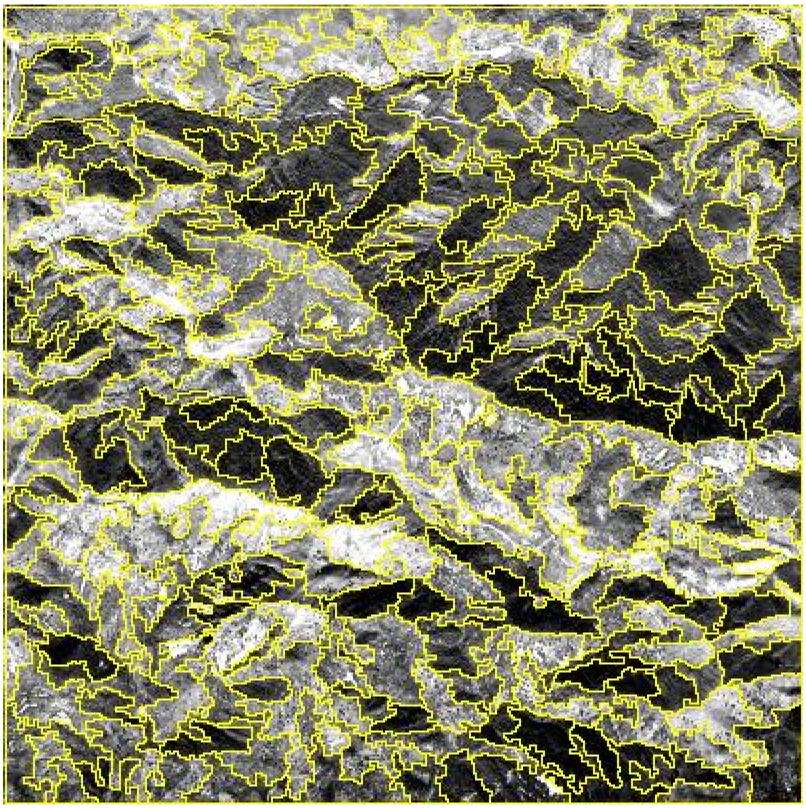


Figure 3-22. Image 2: baseline partition (MMU=2 ha, pixel size=10 m)

3.11. Discussion

Although the method is in all likelihood susceptible of improvement, these preliminary results seem to adapt reasonably well to the spatial structure of the image, at least visually. That is to say that there seems to be some sense of cohesiveness throughout the area enclosed within most of the segments (small gaps excluded, which have been subsumed into larger segments due to the MMU constraint), and contrariwise, some sense of discontinuity across the boundary between each segment and its neighbours. Nevertheless, it should be noted that they evidence some potential problems.

First, the lack of match in relatively uniform areas between partitions obtained with different filtering options, as e.g. between figures 3-5 and 3-6, manifests a high sensitivity of the method to the type and level of smoothing chosen. This is not a surprise, since each filtering scheme will produce a somehow unique simplification of the original image and hence a different watershed partition. The greater the time (no. of iterations) and diffusivity applied, the larger the initial segments and the lower the edge density.

In general, high contrast edges remain the same under all the options. The inconsistencies arise for weak edges, which depending on the particular configuration of the filter will survive the smoothing or not. A similar situation will occur also for human interpreters, whose individual interpretations of low contrasted areas are likely to differ too. Anyway, this problem is common to any segmentation method (see e.g. (Baatz & Schape 2000)). In the case of the baseline method, it could be argued that in order to avoid such inconsistencies, one should simply use the gradient magnitude of the original image. However the output from smoothed versions seems to be better, both visually and quantitatively (lower nRMSE), since smoothing reduces edge intricacy. In any case, it is likely that each interval of the ratio MMU/pixel size requires a different filtering scheme. This issue deserves a specific research that may be addressed in future work.

The problem is that when the method is used for change detection or map updating, slight luminance changes in the new image will produce a very different partition in low-contrast areas. Such changes may have been produced by different illumination, viewing or atmospheric conditions, while patches on the ground may remain unchanged. If such situation corresponds to inner granules (those lying in the inside of polygons and thus not contributing to their boundaries), it does not constitute a problem. But when polygons are separated by weak edges (i.e. when the semantic difference between two adjacent polygons involves only a slight radiometric difference), border granules are likely to change their shape without implying a real change on the ground.

There are two possible ways of tackling spurious change in these areas. The first one would consist in a visual check (of e.g. fresh higher resolution imagery acquired by an unmanned aerial vehicle) to ascertain whether the original boundaries should be substituted by the new ones (when the latter fit better landcover variations observed in the finer image) or otherwise be retained. The second one acknowledges explicitly that boundary displacement between consecutive updates does not necessarily imply a change of landcover type in the affected area. It would give an estimate of the positional reliability (i.e. a probabilistic epsilon band) of each arc of the partition based on its dynamics (i.e. its geodesic saliency under the watershed analogy (Najman & Schmitt 1996), see 3.7). Arcs with low dynamics (weak edges) are more likely to suffer spurious displacement in future images than higher dynamics arcs. Conversely, when strong edges shift in subsequent images, it is probably due to a significant change on the ground.

Nevertheless, it is worth noting that an important exception occurs in hilly terrain. The appearance of shady hillsides depends on solar elevation and azimuth, and then we can have sharp boundaries created by shading that are liable to shift from one image to another. However, they can be easily evaluated through a simulated shaded relief (with a solar position equal to the one of the acquisition time of the image) derived from a DEM, as in figures 3-19 and 3-20. Many granules fit luminance variations in the shaded relief, indicating a strong influence of terrain aspect and slope on the luminance variations of RS images in hilly terrain, especially when solar elevation is low. Notwithstanding the foregoing, this phenomenon is compatible and even helps to identify landcover patches defined as contiguous areas of similar physiognomy and physiography, since significant changes in the latter can be detected for most solar angles provided a DEM is available.

Anyhow, the second way of tackling spurious change is a more economic and honest way to reflect what we really can learn about the territory through the map. It would also help to reduce 'the number of users that believe printed or digital maps to be "true", just as an alarmingly large number of individuals seem to believe anything they read in a formal published source' (Mark & Csillag 1989). That is to say that maps are models, and thus they always contain errors that typically are not distributed uniformly throughout them. Hence users should be aware that the reliability of the information portrayed in maps varies with location. Such sense of varying uncertainty could be conveyed visually by e.g. using increasingly blurred lines for arcs of lower dynamics.

A less flexible alternative would be to work only with arcs of higher dynamics, which are inherently more robust. In this case, the increased certainty of boundary placement has to be balanced with the difficulty of delineating automatically edges that separate transitional areas. Vg compare figures 3-9 (standard watershed) and 3-10 (higher dynamics watershed). In the latter, there are gradations that get encompassed within a larger segment. This is not a problem as long as the gradation involves no change in meaning. For example, the big segment in the left middle of the image covers a good part of the hilltop of a mesa populated with oak trees. In the central part of the mesa, water availability is lower and insolation is higher, leading to a more sparse distribution of trees. Being the trees from the same species, such difference is not relevant for the MFE, therefore all the hilltop is included in the same polygon. But in many other cases gradation may be significant, and new boundaries separating semantically different regions that blend into each other would have to be drawn by other means.

Another problem of the current implementation of the baseline method is the dependence of the output on the update frequency during the merging process of the labelled image representing the partition. This is illustrated in figures 3-7 and 3-8. The baseline partition (MMU=5 ha) was computed for image 1a and for a central 200x200 pixel subset of it. The red lines correspond to arcs that are not present in the subset partition. Most of them are close to the subset border, indicating that the lack of context (given by the areas surrounding the subset) is a factor contributing to these differences. But also there are non-conforming arcs in the central part of the subset. Then the problem is that the larger partition was updated at different iterations than the one of the subset. Every time a new partition is updated by scanning the labelled image instead of the more economic alternative of recomputing the attributes from the lists, the size and signature of current segments are slightly modified due to the inclusion of subsumed watershed pixels. As a result, the merging sequence in the next iteration may be different from the one that would have been obtained by recomputing the attributes from the list. A safe solution would be to scan the labelled partition image after each iteration, but it would be too slow. In any case, this is a problem affecting not the baseline method itself but the current implementation, and as such will be addressed in future work. In particular, it has to be resolved before the method can be applied to big images (say > 2000x2000 pixels) requiring tiled processing.

Other inconsistencies appear when the baseline partition is derived for the same scene using different band combinations (figures 3-13, 3-14 and 3-15). Again, this was to be expected, since different inputs produce differing outputs no matter what algorithm is used. However, a higher degree of match would be desirable for the method to be robust. In fact, the shape of high contrasted regions (like the burn scar in the top right angle of the image) does not suffer significant change, and the same can be said for strong edges. But weak edges are combined in very different ways, producing disparate regions. Nevertheless, none of the partitions produce a visual impression of a 'bad segmented' image, since for most segments, there seems to be some sense of cohesiveness throughout the area enclosed within the segment, and contrariwise, some sense of discontinuity across the boundary between the segment and its neighbours. As a matter of fact, similar inconsistencies would show up if the scene were interpreted by the same person using different colour composites.

In the end, such inconsistencies arise because, as stated in 2.2.1, the representations of reality are manifold, and in many situations, none of them can be said to be strictly preferred to the others. For example, it could be chosen the one with the lowest nRMSE, but this error estimate is not only dependent on the input data set but on the definition of error. Perhaps for this reason, visual check remains the basic evaluation procedure for newly developed algorithms, although there are some empirical methods (see (Zhang 1996) for a review) that try to mitigate the inevitable subjectiveness of the evaluation. Besides, error reduction may be only one of several conflicting goals in landcover mapping. In short, there is no single correct patch hierarchy fitting a given landscape, since hierarchies and the maps that represent them are human constructs (Wiens 1995). Each individual and institution has different interests, conceptions and methods, therefore they may hold different views of the same reality. In spite of this, there should be a good correspondence between model and reality, so that users operating on the model obtain roughly the same results than users operating on reality, providing both groups share the same interest and conceptions than the producer of the map. Since users and operations are manifold, the issue of producing the best model requires a deeper analysis under the framework provided by multiobjective decision-making theory (see e.g. (Chankong & Haimes 1983)). Such analysis is beyond the scope of this thesis, although it should be addressed in future research in order for the method to constitute an operational standard in image analysis for landcover mapping.

An example of the weaknesses of the spectrometric approach is shown in figure 3-18. The fifteen spectral classes defined by the K-means algorithm did not result in meaningful information classes unless they were grouped into very general categories. Even after aggregation, some combined classes still reveal some inconsistencies. Vg the burn scar of the top right angle of the image was classified as a sparse woodland, whereas in the field check it was found to be covered by rockroses. In other cases, the inconsistency arises because of the internal heterogeneity of MFE polygons. Vg the central part the fore-mentioned mesa has been correctly classified as a sparse woodland, although in the MFE it has been included within a forest polygon for reasons discussed before. Also, the cliff facing the South hillside of the mesa has been classified as a sparse woodland, whereas it is populated by thicket. The reason in this case is the lower radiance coming from the cliff, because of its aspect and slope. The conclusion is that a sound classification cannot rely solely on radiometric signatures, there is a wealth of relational features that should be taken into account additionally. Such features, many of them related to physiographic attributes, can only be exploited by a rule-based scheme that may be formalized into e.g. a semantic network.

However, a radiometric classification of granules (or even better, of the blobs compounding each granule, since they are more homogeneous radiometrically; in this case, the class allocated to the granule can be e.g. the most frequently found within it) may serve as a preliminary guide to detect incongruent (having a different appearance) zones within each polygon of a to-be-updated map. Such incongruencies can be evaluated by a set of logic rules. Those that cannot be resolved by the system would be marked for inspection as proposed in 2.6.2.2. The example of figure 3-18 points towards this possibility. Most MFE polygons are compounded by granules belonging to a prevailing class. Granules non conforming to that class may be assessed according to the label of the polygon in which they are encompassed. After evaluation, polygon boundaries can be redrawn using the outer edges of the granules overlapping with it and conforming to its label. The new shape of the MFE layer would be similar to the one shown in figure 3-16, except for the fact that no evaluation was carried out in this case, the only criterion was to allocate granules to polygons according to their geographic overlap.

Finally, an interesting phenomenon related to the fractal nature of geographic boundaries (Goodchild & Mark 1987) is shown in figures 3-21 and 3-22. The baseline partition (MMU=2 ha) was derived for image 2 using i) the original pixel size (1m) and ii) a pixel size of 10 m. The most conspicuous difference between them is the higher edge complexity of the 1 m partition, suggesting that the length of the boundary enclosing any given patch increases indefinitely as the resolution increases (and therefore as the interval between vertices is reduced). This behaviour is in sharp contrast to the one of mathematically differentiable contours when approximated with a polygonal path, whose length reaches a finite limit as the vertex interval approaches zero. Such paradox (that the measurement of length increases with increased accuracy), first reported by Steinhaus (Steinhaus 1960), was solved and explained by a power law (in which the exponent is the fractal dimension) in Mandelbrot's (Mandelbrot 1982) Fractal Theory. A typical fractal curve is shown in figure 3- .

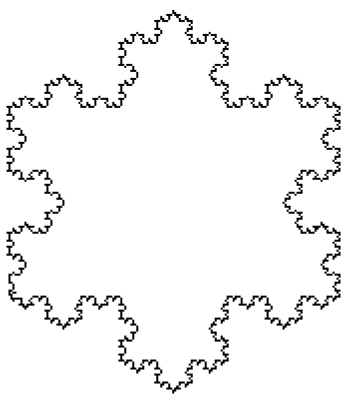


Figure 3-23. Koch's snowflake

The practical implications of this phenomenon are twofold. One the one hand, given a MMU size and an intended scale of representation, there should be an optimum resolution balancing edge simplicity and accuracy. It has been argued (Goodchild and Mark 1987) that such optimum could correspond to the resolution that yields the highest fractal dimension within the range of feasible resolutions. Lacking a deeper applied study, a tentative rule of thumb could be to select a pixel size between 25 and 250 times smaller than the MMU size. This means e.g. that for a MMU of 2 ha, pixel sizes between 10 and 30 m would be suitable. On the other hand, the fractal nature of patch boundaries requires specific algorithms for displaying them at various scales. This is the province of cartographic generalisation (see (Brassel & Weibel 1988) for a review), and as such it is beyond the scope of this thesis. However, it is worth noting that 'despite more than 25 years of concerted effort, fully automated generalization is still a dream, albeit one within reach' (Shuurman 1999).

CHAPTER 4

Conclusions

4.1. Conclusions

The main conclusions of this thesis are summarised below:

1) Information comes ultimately from physical order. Order implies differences, that is, non-uniform distributions of matter/energy. Certain organisms –users- having sensors can take advantage of these differences for cognitive purposes, provided they can detect them. From all the set of detectable differences (*latent information*), only the relevant ones are used by those individuals to construct a handy representation (*structural information*) of the sensed scene, that has to be formalised into a model in order to be communicated. Hence the information portrayed in any given map can be generically defined as a formal representation of a territory whose meaning can be agreed by a community of users.

2) Information is not extracted from images but produced during the analysis. The result of the analysis is the imposition of a simpler, meaningful structure upon the intricate original one, that is, a formal representation (a model) of the structure of the image. This model is dependent not only on the data alone, but also on the definition of the type of objects we want to foreground and on how detailed the representation is intended to be. Different choices during the process of analysis will yield different representations (and hence different information) from the same data set.

3) Conventional quantitative methods of image classification can be conceptualised as nested into the framework of two general accepted approaches or *paradigms*. The wider paradigm, called the *spectrometric approach*, exploits the structure of the multidimensional data space in order to discriminate between the waveforms, or *signatures*, associated to each landcover class. Signatures are formed, in the view supplied by the narrower *pixel-based* paradigm, by individual pixels, which are allocated to the user-defined classes with the aid of discriminant functions.

4) The two basic assumptions underlying the spectrometric approach (**A:** that the piece of terrain from which the measurement is drawn is big enough to include a sufficient number of elements producing the typical response of a landcover class; and **B:** that each landcover class shows a negligible degree of overlap in the data space with the other classes) cannot be fulfilled simultaneously if the basic units of the analysis consist in individual pixels.

5) In order for the spectrometric approach to be successful, there should be a prevailing landcover class in each cluster of the data space. Since there are many ways in which landcover patches can be conceptualised and individuated, it is too optimistic to expect that clusters keep a one-to-one correspondence to the geographic concepts that we use to divide up the landscape into meaningful chunks. On the one hand, classes are formed by family resemblances following a set of prototypical instances. Therefore the mathematical definition of classes is an ill-posed problem. On the other hand, the surroundings (and even inner gaps) of geographic objects may be significant for classification. Therefore a sound classification cannot rely solely on radiometric signatures, there is a wealth of relational features that should be taken into account additionally.

6) The spectrometric approach considers class-concepts as *mass nouns* referring to homogeneous materials. Such view cannot account for the hierarchically structured heterogeneity of landscapes. It is also in contrast to the hierachical patch model underlying modern landscape ecology. The latter conceives the piece of terrain enclosed by a polygon as a referent to a *count noun,* i.e. as an instance of some type of geographic object of a particular level within the hierarchy. Therefore an object-oriented approach (compatible with the landscape conception of the target model) to classification is more appropriate than the spectrometric approach.

7) The qualitative space represented by landcover maps is a partition of the quantitative space defined by the geographic fields depicting the variation of relevant biophysical attributes throughout the territory. The boundaries separating polygons correspond to zones where some attribute changes abruptly and hence can be interpreted as qualitative discontinuities. The latter are explained mathematically by René Thom's (1975) catastrophe theory. This theory can be applied successfully to landcover mapping by establishing an analogy between stable attractors and the local minima of a gradient magnitude image representing the overall variation of the relevant attributes within the territory. Then the basins of attraction defining the morphology of the phenomenon are the catchment basins of the watershed transform of that gradient magnitude image.

8) RS images can be taken as a surrogate of the geographic fields associated to the biophysical attributes of landcover. However, the correspondence between the value of the images at a given geographic point and the one of landcover attributes is variable, difficult to

determine, and dependent on the spatio-temporal scale of observation. These shortcomings can be tackled if, instead of focusing on point-wise observations, the spatial variation of the images is studied. The basic premise of the approach is that the overall spatial variation of the latter coincides to a great deal with the one of the attribute fields. By applying the above analogy to the image, it is assumed that each structural-functional unit of the image corresponds to structural-functional unit in the landscape, i.e. that each blob of the image is a patch of the landscape, where a blob is the catchment basin of a gradient minimum.

9) The hierarchical nature of the landscape, together with the great variability in size and appearance of the patches within each level of the hierarchy, make untenable the hypothesis that, given an image with a fixed pixel size, each blob corresponds to a patch of the same hierarchic level. If on the one hand we set the basic level at patches defined as a contiguous area of similar dominant species, physiognomy and physiography, and on the other hand we use a high resolution image, most blobs will have to be aggregated into larger units. Then two additional assumptions have to be made: i) if blobs correspond to significant patches of this level or higher, their projection onto the ground must exceed a given size; and ii) if two adjacent blobs are radiometrically similar, they are semantically similar too. The second premise provides a rule to merge adjacent blobs, while the first one defines the merging stop criterion.

10) Objects are class instances, and in the case of landcover they are patches that qualify as referents to the concepts used by the classification scheme of the map to divide up the landscape into landcover types. Then, in order for a patch to become a geographic object, it should have enough extension as to deserve inclusion in the map as an individual entity. This threshold is defined through the MMU size, so that patches below this size can never become instances of classes at that level of generalisation. Therefore, under the object-oriented approach, it is required that the spatial units subject to classification exceed the MMU size. Hence the first goal of object-oriented classification is to partition the territory into a set of basic mappable zones, or *granules*, exceeding the MMU size.

11) Such partition can be achieved by i) applying to a RS image ortho-image a non-linear diffusion filter that gets rid of superfluous gradient minima created by texture and/or noise; ii) detecting the blobs of the filtered image via gradient watersheds; and iii) merging adjacent blobs according to their radiometric similarity in the original image until they reach the MMU size.

12) Preliminary results based on this procedure seem to adapt reasonably well to the spatial structure of the image, at least visually. There seems to be some sense of cohesiveness throughout the area enclosed within most of the granules, and contrariwise, some sense of discontinuity across the boundary between each granule and its neighbours. However, the results reveal some potential problems:

i) low degree of match in low contrasted areas between partitions obtained with different filtering options, indicating a lack of robustness in those regions.

ii) dependence of the output on the update frequency during the merging process of the labelled image representing the partition, that has to be solved before the method can be applied to big images requiring tiled processing.

iii) fractal effects (that the length of the boundary enclosing any given patch increases indefinitely as the resolution increases) in the output boundaries that may difficult arc simplification during vectorisation.

4.2. Future work

This thesis provides a conceptual framework and an automated method on which to base object-oriented classification of RS images for landcover mapping. The ideas and algorithms it contains are susceptible of refinement, and subsequent areas of research can be summarized in three items, related respectively to the R-model, the baseline method, and the classification of granules.

4.2.1. Refinement of the R-model

The realistic model in which the conceptual framework presented in this thesis has been formalized is itself susceptible of further formalization into a logico-mathematical set of axioms, definitions and properties/relations. The theory of granular partitions (Bittner and Smith 2001) is good starting point for such effort. In addition, the three hypotheses in which the R-model is based deserve a deeper analysis and even empirical testing. The latter could be carried out in a place where extensive ground inventory data are available together with coetaneous RS imagery. A good candidate is the Duke Forest in North Carolina (http://taxodium.env.duke.edu/forest/). This research would require a previous investigation on the definition of a suitable semantic dissimilarity measure with which to estimate the overall gradient magnitude image of the idealistic fields depicting the spatial variation of biophysical attributes over the test site. The effect on the coincidence hypothesis of differing areal supports between RS images and biophysical attributes should also be evaluated.

4.2.2. Improvement of the baseline method

In order to for the baseline method to become an operational standard in image analysis for landcover mapping, it needs further testing and optimization. Several filtering methods should be compared regarding their output and performance. In particular, it is likely that a different filtering scheme is best fitted for each interval of the ratio MMU/pixel size. Probably diffusivity should increase with this ratio, and a research proposing suitable parameter setting for each ratio interval is desirable.

It would also be desirable to test another dissimilarity measures. For example, the simple Euclidean distance produces better results in terms of RMSE, but visually NVD results are better. However, in hyperspectral images it may be advisable to use the former, since chrominance in high dimensional data spaces may lose sense.

There are another ways of computing the gradient magnitude image that could also be evaluated. The one chosen is the simplest one, but e.g. it is not invariant to image rotation. However, I believe that the impact of different computations on the final result is negligible, but in any case it should be tested.

The simplification of the watershed partition by using edge dynamics (which was discarded a priori in the current implementation) should also be considered for evaluation in subsequent implementations. Weak watershed arcs are very sensitive to slight luminance changes in new images. If these changes are due to different illumination conditions rather than to changes on the ground, the resulting configuration of the new partition may be misleading in low contrast areas. Therefore it may be preferable to get rid of weak edges prior to the merging step. The result would gain robustness at the expense of i) some arbitrariness in the selection of a suitable threshold; ii) a higher disparity in size of the resulting granules; and iii) an increased difficulty in dealing with transitional areas.

The SCRM algorithm is also susceptible of improvement. In particular, a procedure for the updating of the partition avoiding the inconsistencies (due to the preclusion of physical update in normal iterations) of the current implementation is required for tiled processing. A possible solution is to convert the watershed partition to a polyline vector layer in which the attributes of each arc consist of the mean value of its pixels in each band, and the label of the catchment basins it separates. Then the size and the mean value of each merger could be precisely computed with the aid of the layer database without having to access the labelled image.

Finally, a line generalisation algorithm dealing adequately with the fractal nature of the boundaries of the baseline partition should be found and implemented into the method. One possible option is to place the vertices compounding the vector layer into a scalar hierarchy in a similar fashion than with regions, so that each coarser scale of representation has simpler edges as well as fewer regions. A related issue is the selection of the optimum working pixel size.

4.2.3. Classification of granules

The classification of granules to form the final mapping units requires, on the one hand, the definition of a set of surrogate (non biophysical) attributes in which to base the classification, and on the other, a procedure to carry it out. In 2.6.2.2 there are some suggestions on how to perform both tasks that can be used as guidelines for future work. The procedure to be designed could also be inspired or benchmarked by the hierarchical semantic network of image objects used by the e-cognition software (http://www.definiens-imaging.com), in which fine-scale structures (blobs in our case) are sub-objects of coarser objects (granules in our case). The particular arrangement and properties of the blobs within a granule can then be taken as structural attributes that in turn may be used to define quantitatively each landcover class. Another interesting idea from this software is the use of fuzzy membership functions to enable comparison between quantitative and relational attributes. These functions are defined separately for each class and attribute, and indicate the degree of membership (a value between 0 and 1) to a given class that can be expected for a granule that show a particular value of this attribute. Once defined, such functions could be used to construct the concordance, discordance and indifference subsets in the ELECTRE method.

REFERENCES

Abramson, S.B. and Schowengerdt, R. A. (1993) Evaluation of edge-preserving smoothing filters for digital image mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* 48 (23): 2-17.

Ahl, V. and Allen, T. F. H. (1996) *Hierarchy Theory: A Vision, Vocabulary, and Epistemology*. Columbia University Press, New York.

Ahlcrona, E. (1995) CORINE Land Cover: A pilot project in Sweden. In: *Sensors and Environmental Applications of Remote Sensing* (ed Askne, J.), pp. 19-22. Balkema, Rotterdam.

Alker, H.R. (1969) A Typology of Ecological Fallacies. In: *Quantitative Ecological Analysis in the Social Sciences* (eds Dogan,M. and Rokkan,S.), MIT Press.

Armitage,R.P., Weaver, R. E., and Kent, M. (2000) Remote sensing of semi-natural upland vegetation: the relationship between species composition and spectral response. In: *Vegetation Mapping: From Patch to Planet* (eds Millington,A. and Alexander,R.), John Wiley and sons.

Asrar, G. (1989) *Theory and Applications of Optical Remote Sensing*. Wiley, New York.

Baatz, M. and Schape, A. (2000) Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: *Angewandte Geographische Informationsverarbeitung XII* (eds Strobl,J. and Blaschke,T.), pp. 12-23. Wichmann-Verlag, Heidelberg.

Baraldi, A. and Parmiggiani, F. (1994) A Nagao-Matsuyama approach to high-resolution satellite image classification. *IEEE Trans.Geosci.Remote Sensing* 32 (4): 749-758.

Baraldi, A. and Parmiggiani, F. (1996) Single linkage region growing algorithms based on the vector degree of match. *IEEE Transactions on Geoscience and Remote Sensing* 34 (1): 137-148.

Barham, J. (1996) A dynamical model for the meaning of information. *Biosystems* 38: 235-241.

Bateson, G. (1972) *Steps to an ecology of mind*. Ballatine, New York.

Beaulieu,J.M. and Goldberg, M. (1989) Hierarchy in picture segmentation: A stepwise optimisation approach. *IEEE Transactions on Pattern Analysis and Machine Inteligence* 11: 150-163.

Benitez, J.M., Castro, J. L., and Requena, I. (1997) Are Artificial Neural Networks Black Boxes? *IEEE Transactions on Neural Networks* 8 (5): 1156-1164.

Bennett, B. (2001) What is a Forest? on the vagueness of certain geographic concepts. *Topoi* 20 (2): 189-201.

Beucher, S. and Lantuejoul, C. (1979) Use of watersheds in contour detection. Proceedings, International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation . Rennes.
Ref Type: Conference Proceeding

Bittner, T. (1999) On Ontology and Epistemology of Rough Location. Proceedings of COSIT 1999, pp. 433-448.

Bittner, T. and Smith, B. (2001a) A Taxonomy of Granular Partitions. In: *Foundations of Geographic Information Science, Proceedings of COSIT 2001* (ed Montello, D.), pp. 28-43. Springer.

Bittner, T. and Smith, B. (2001b) Vagueness and Granular Partitions. In: *Formal Ontology and Information Systems* (eds Welty, C. and Smith, B.), pp. 309-321. ACM Press, New York.

Bittner, T. and Winter, S. (1999) On Ontology in Image Analysis. In: *Integrated Spatial Databases: Digital Images and GIS* (eds Agouris,P. and Stefanidis,A.), pp. 168-194. Springer-Verlag.

Blaschke, T. and Hay, G. (2001) Object-oriented image analysis and scale-space: Theory and methods for modeling and evaluating multi-scale landscape structure. *International Archives of Photogrammetry and Remote Sensing* 34: 22-29.

Booch, G. (1991) *Object-Oriented Design with Applications*. The Benjamin/Cummings Publishing Company Inc., Redwood City.

Brachman, R.J. (1977) What's in a concept: Structural foundations for semantic networks. *International Journal of Man-Machine Studies* 9 (127): 152.

Brassel, K.E. and Weibel, R. (1988) A Review and Conceptual Framework of Automated Map Generalization. *International Journal of Geographical Information Systems* 2 (3): 229-244.

Bronowski, J. (1970) New concepts in the evolution of complexity: Stratified stability and unbounded plans. *Synthese* 21: 228-246.

Bruegger, B.P. (1994) *Spatial theory for the integration of resolution-limited data.* Ph.D. Thesis, University of Maine.

Buttner, G., Hajos, T., and Korandi, T. (1989) Improvements to the Effectiveness of Supervised Training Procedures. *International Journal of Remote Sensing* 10 (6): 1005-1013.

Campbell, D.T. (1988) *Methodology and epistemology for social science: selected papers*. University of Chicago Press.

Casati, R., Smith, B., and Varzi, A. (1998) Ontological Tools for Geographic Representation. In: *Formal Ontology in Information Systems* (ed Guarino, N.), pp. 77-85. IOS Press.

Chandler, D. (2001) *Semiotics: The Basics*. Taylor and Francis Group.

Chankong, V. and Haimes, Y. (1983) *Multiobjective Decision Making Theory and Methodology*. North-Holland.

Chmielecki , A. (1998) What is information? 20th World Conference of Philosophy. http://www.bu.edu/wcp/Papers/Cogn/CognChmi.htm

Chrisman, N.R. (1991) The Error Component in Spatial Data. In: *Geographical Information Systems: Principles and Applications* (eds Maguire,D.J., Goodchild,M.F., and Rhind,D.W.), Longman Scientific and Technical.

Cohn, A.G. and Gotts, N. M. (1996b) The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries. In: *Geographic Objects with Indeterminate Boundaries* (eds Burrough, P.A. and Frank, A.U.), pp. 171-188. Taylor and Francis Inc.

Corning, P.A. (2001) Control Information: The Missing Element in Norbert Wiener's Cybernetic Paradigm? *Kybernetes* 30: 1272-1288.

Cox, I.J. (1993) A Review of Statistical Data Association Techniques for Motion Correspondence. *International Journal of Computer Vision* 10 (1): 53-66.

Cressie, N.A.C. (1996) Change of support and the modifiable areal unit problem. *Geographical Systems* 3: 159-180.

Cross, A.M., Mason, D. C., and Dury, S. J. (1988) Segmentation of remotely-sensed images by a split-and-merge process. *International Journal of Remote Sensing* 9 (8): 1329-1345.

Cufi,X., Muñoz, X., Freixenet, J., and Martí, J. (2002) A Review on Image Segmentation Techniques Integrating Region and Boundary Information. In: *Advances in Imaging and Electron Physics, vol. 120* (ed Hawkes,P.W.), pp. 1-39. Academic Press.

Curran,P.J. (1985) *Principles of remote sensing*. Longman Group Limited, London.

Curran,P.J. (1988) The semivariogram in remote sensing: an introduction. *Remote Sensing of Environment* 24: 493-507.

Davis,F.W. and Simonett, D. S. (1991) GIS and remote sensing. In: *Geographic Information Systems: Principles and Applications* (eds Maguire, D.J., Goodchild, M.F., and Rhind, D.), pp. 191-210. Longman, Essex, England.

Derin, H. and Cole, W. S. (1986) Segmentation of textured images using Gibbs random fields. *Computer Vision Graphics and Image Processing* 35: 72-98.

Egmont-Petersen, M., de Ridder, D., and Handels, H. (2002) Image processing with neural networks - a review. *Pattern Recognition* 35 (10): 2279-2301.

Ehlers, M., Edwars, G., and Bedard, Y. (1989) Integration of remote sensing with geographic information systems: a necessary evolution. *Photogrammetric Engineering and Remote Sensing* 55: 1619-1627.

FAO (1989) Principles of radar imagery. Remote Sensing series, no. 62.

Ferro, C.J. (1998) *Scale and texture in digital image classification.* M.Sc. Thesis, West Virginia University.

Fjørtoft, R., Lopès, A., Marthon, P., and Cubero, E. (1998) An optimal multiedge detector for SAR image segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 36 (3): 793-802.

Fotheringham, A.S. and Wong, D. W. S. (1989) The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* 23: 1025-1044.

Frank, A.U. (2001) Tiers of ontology and consistency constraints in geographic information systems. *International Journal of Geographical Information Science* 15 (7): 667-678.

Frank, A.U. (2002) Ontology for spatio-temporal databases. In: *Spatiotemporal Databases: The Chorochronos Approach.* Springer-Verlag.

Frank, A.U. and Timpf, S. (1994) Multiple Representations for cartographic objects in a multi-scale tree - an intelligent graphical zoom. *Computers and Graphics: Special Issue on Modelling and Visualization of Spatial Data in GIS* 18 (6): 823-829.

Frank,A.U., Volta, G. S., and Mcgranaghan, M. (1997) Formalization of families of categorical coverages. *International Journal of Geographical Information Science* 11 (3): 215-231.

Frege,G. (1892) On Sense and Reference. In: *Translations from the Philosophical Writings of Gottlob Frege (1960)* Blackwell, Oxford.

Fukunaga,K. (1972) *Introduction to Statistical Pattern Recognition*. Academic Press, New York.

Gauch, J.M. (1999) Image segmentation and analysis via multiscale gradient watershed hierarchies. *IEEE Transactions on Image Processing* 8 (1): 69-79.

Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Goldberg, M., Goodenough, D. G., and Plunkett, G. W. (1988) A Knowledge-Based Approach for Evaluating Forestry Map Congruency with Remotely Sensed Imagery. *Philosophical Transactions of the Royal Society of London* A324: 447-456.

Goodchild, M.F. (1992) Geographic data modeling. *Computers and Geosciences* 18 (4): 401-408.

Goodchild, M.F. (1994) Integrating GIS and remote sensing for vegetation analysis and modelling:methodological issues. *Journal of Vegetation Science* 5: 615-626.

Goodchild, M.F. and Mark, D. M. (1987) The fractal nature of geographic phenomena. *Annals of the Association of American Geographers* 77 (2): 265-278.

Goodchild, M.F. and Quatrocci, D. A. (1997) Scale, Multiscaling, Remote Sensing, and GIS. In: *Scale in Remote Sensing and GIS* (eds Quatrocci,D.A. and Goodchild,M.F.), pp. 1-11. Lewis.

Gödel, K. (1931) Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38: 173-198.

Graetz, R.D. (1990) Remote Sensing of terrestrial ecosystem structure: an ecologyst's pragmatic view. In: *Remote Sensing of Biosphere functioning* (eds Hobbs,R.J. and Mooney,H.A.), Springer-Verlag, New York.

Green, D.R. and Hartley, S. (2000c) Integrating photointerpretation and GIS for vegetation mapping: some issues of error. In: *Vegetation Mapping: From Patch to Planet* (eds Millington, A. and Alexander, R.), pp. 103-134. John Wiley and sons.

Haar Romeny, B. T. (1997) Introduction to Scale-Space theory: multiscale geometric image analysis. Utrecht, the Netherlands, Springer Verlag. First International Conference on Scale-Space theory.

Hagner, O. (1990) Computer-aided forest stand delineation and inventory based on satellite remote sensing. In: *Proceedings of the SNS/IUFRO workshop on the usability of remote sensing for forest inventory and planning* pp. 94-105. Swedish University of Agricultural Sciences, Remote Sensing Laboratory,Report 4, Umeå.

Haralick, R.M. (1979) Statistical and structural approaches to texture. *Proceedings of the IEEE* 67 (5): 786-804.

Haralick, R.M. and Saphiro, L. G. (1985) Survey: image segmentation techniques. *Computer Vision Graphics and Image Processing* 29: 100-132.

Haris, K., Efstratiadis, S. N., and Katsaggelos, A. K. (1998) Hybrid Image Segmentation Using Watersheds and Fast Region Merging. *IEEE Transactions on Image Processing* 7 (12): 1684-1699.

Hartigan, J.A. and Wong, M. A. (1979) A K-Means Clustering Algorithm. *Applied Statistics* 28 (1): 100-108.

Hay, G.J., Marceau, D. J., Dube, P., and Bouchard, A. (2001) A Multiscale Framework for Landscape Analysis: Object-Specific Analysis and Upscaling. *Landscape Ecology* 16 (6): 471-490.

Hobbs, S. E., Ang, W., and Seynat, C. (1998) Wind and Rain Effects on SAR Backscatter from Crops. ESA SP-441: 185-190.

Hoffmeyer, J. (1997) *Signs of Meaning in the Universe*. Indiana University Press.

Holbo, H.R. and Luvall, J. C. (1989d) Modeling surface temperature distributions in forest landscapes. *Remote Sensing of Environment* 27: 11-24.

Honeycutt, D. M. (1987) Epsilon bands bases on probability. Proceedings, Autocarto-8.

Horton, R. (1982) Tradition and Modernity Revisited. In: *Rationality and Relativism* (eds Hollis, M. and Lukes, S.), pp. 201-260. Blackwell, Oxford.

Hsieh, P. F. and Lee, L. C. (2000) Effect of spatial resolution on classification error in Remote Sensing. IGARSS 2000.

Huang, T. (1969) Per Field Classifier For Agricultural Applications. LARS Technical Note 060569: 1-14. Purdue University.

Jackway, P.T. (1996) Gradient Watersheds in Morphological Scale-Space. *IEEE Transactions on Image Processing* 5 (6): 913-921.

Janssen, L.L.F. and van der Wel, F. J. M. (1994) Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing* 60 (4): 419-426.

Jelinski, D.E. and Wu, J. (1996) The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology* 11: 129-140.

Ji, S. and Park, H. W. (1998) Image segmentation of color image based on region coherency. *Proceedings of ICIP 98'* 1: 80-83.

Jupp, D.L., Strahler, A. H., and Woodcock C. (1989) Autocorrelation and regularization in digital images: II. Simple image models. *IEEE Transactions on Geosciences and Remote Sensing* 27: 247-256.

Kaiser, P.K. and Boynton, R. M. (1996) *Human Color Vision*. Optical Society of America, Washington, DC.

Kartikeyan, B., Sarkar, A., and Majumder, K. L. (1998) A Segmentation Approach to Classification of Remote Sensing Imagery. *International Journal of Remote Sensing* 19 (9): 1695-1709.

Kettig, R.L. and Landgrebe, D. A. (1976) Classification of Multispectral Image Data by Extraction and Classification of Homogenous Objects. *IEEE Transactions on Geoscience and Remote Sensing* GE-14: 19-26.

Klenk,J., Binnig, G., and Schmidt, G. (2000) Handling Complexity with Self-Organizing Fractal Semantic Networks. *Emergence* 2 (4): 151-162.

Koch, K.R. (1988) *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer, Berlin.

Koestler, A. (1967b) *The Ghost in the Machine* . Henry Regnery.

Korzybski, A. (1933) *Science and Sanity*. International non Aristotelian Library Publishing Co.

Kuchler, A.W. (1967) *Vegetation Mapping*. The Ronald Press Company, New York.

Kuhn,T.S. (1962) *The Structure of Scientific Revolutions*. The Chicago University Press.

Kuijper, A. and Florack, L. (2001) The application of catastrophe theory to image analysis. Report UU-CS-2001-23. University of Utrecht.

Lakoff, G. (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

Landgrebe, D. (1999) Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data. In: *Information Processing for Remote Sensing* (ed Chen,C.H.), The World Scientific Publishing Co., New Jersey.

Landgrebe, D.A. (1980) The development of a spectral-spatial classifier for earth observational data. *Pattern Recognition* 12: 165-175.

Legendre, L. and Legendre, P. (1998) *Numerical Ecology*, 2nd English edn. Elselvier, Amsterdam.

Lindeberg, T. (1993) Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus of attention. *International Journal of Computer Vision* 11 (3): 283-318.

Lindeberg, T. (1994) Scale-space theory: a basic tool for analysing structures at different scales. *Journal of Applied Statisitics* 21 (2): 225-270.

Lins, K. (1994) Requirements analysis results for land cover and land use data. *US Geological Survey* March

Lobo, A. (1997) Image segmentation and discriminant analysis for the identification of land cover units in ecology. *IEEE Transactions on Geoscience and Remote Sensing* 35 (5): 1136-1145.

Mandelbrot, B.B. (1982) *The Fractal Geometry of Nature*. W.H. Freeman, New York.

Marceau, D.J. (1999) The scale issue in social and natural sciences. *Canadian Journal of Remote Sensing* 25 (4): 347-356.

Marceau, D.J., Howarth, P. J., Dubois, J. M., and Gratton, D. J. (1990) Evaluation of the Grey-Level Co-occurrence Matrix Method for Land-Cover Classification Using SPOT Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 28 (4): 513-519.

Marceau, D.J., Howarth, P. J., and Gratton, D. J. (1994) Remote Sensing and the Measurement of Geographical Entities in a Forested Environment. Part 1: The Scale and Spatial Aggregation Problem. *Remote Sensing of Environment* 49 (2): 93-104.

Mark, D.M. and Csillag, F. (1989) The nature of boundaries on area-class maps. *Cartographica* 21: 65-78.

Mark, D.M., Smith, B., and Tversky, B. (1999) Ontology and geographic objects: an empirical study of cognitive categorization. *Springer lecture notes in Computer Science (COSIT '99)*: 283-298.

Markham, B. L. and Townshend, J. R. G. (1981) Land cover classification accuracy as a function of sensor spatial resolution. Proc.15th Int.Symp.Remote Senssing Environ. 1075. Ann Arbor.

Marr, D. (1982b) *Vision*. Freeman Publishers.

Matheron, G. (1971) *The theory of regionalized variables and its applications*. Paris School of Mines publication.

Matheron, G. (1983) Isofactorial methods and Change of Support. Proceedings 2nd NATO ASI Geostatistics for Natural Resources Characterization. Dordrecht, Holland.

Mausel,P.W., Kramber, W. J., and Lee, J. K. (1990) Optimum Band Selection for Supervised Classification of Multispectral Data. *Photogrammetric Engineering and Remote Sensing* 56 (1): 55-60.

Maxwell, J.C. (1870) On dales and hills. *The London, Edinburgh and Dublin Philosophical Magazine* 40: 421-425.

Meyer, F. (2001) An overview of morphological segmentation. *International Journal of Pattern Recognition and Artificial Intelligence* 15 (7): 1089-1118.

Millington, A. and Alexander, R. (2000) Vegetation Mapping in the Last Three Decades of the Twentieth Century. In: *Vegetation Mapping: From Patch to Planet* (eds Millington, A. and Alexander, R.), pp. 321-331. John Wiley and sons.

Moller-Jensen, L. (1990) Knowledge-based classification of an urban area using texture and context in Landsat-TM imagery. *Photogrammetric Engineering and Remote Sensing* 56 (6): 899-904.

Moore, G.E. (1959) A Defence of Common Sense. In: *Philosophical Papers (Moore, G. E.)* pp. 60-88. George Allen and Unwin, London.

Muenzinger,K.F. (1942) *The Psychology of Behavior*. Harper, New York.

Nagao, M. and Matsuyama, T. (1979) Edge preserving smoothing. *Computer Graphics and Image Processing* 9: 394-407.

Najman,L. and Schmitt, M. (1996) Geodesic Saliency of Watershed Contours and Hierarchical Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (12): 1163-1173.

Narayanan, R. M., Sankaravadivelu, T. S., and Reichenbach, S. E. (2000) Dependence of image information content on gray-scale resolution. IEEE. IGARSS 2000.

Nazif, A.M. and Levine, M. D. (1984) Low level image segmentation: An expert system. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (5): 555-577.

Olsen, O.F. and Nielsen, M. (1997) Multi-Scale Gradient Magnitude Watershed Segmentation. *Lectures Notes in Computer Science* 1310: 6-13.

Openshaw, S. (1978) An optimal zoning approach to the study of spatially aggregated data. In: *Spatial Representation and Spatial Interaction* (eds Masser,I. and Brown,P.), Martinus Nijhoff, Leiden.

Openshaw, S. (1984) The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography (CATMOG) (38). Norwich Geobooks.

Pal, N.R. and Pal, S. K. (1993) A review on image segmentation techniques. *Pattern Recognition* 26 (9): 1277-1294.

Paltridge, G.W. (1978a) The steady-state format of global climate. *Quarterly Journal of the Royal Meteorological Society* 104: 927-945.

Pattee, H.H. (1973) *Hierarchy Theory: The Challenge of Complex Systems*. George Braziller, New York.

Pawlak, Z. (1982) Rough sets. *International Journal of Computers and Information Science* 11 (341): 356.

Perkal, J. (1966) On the length of empirical curves. (Discussion Paper 10). Ann Arbor. Inter-University Community of Mathematical Geographers.

Perona, P. and Malik, J. (1990) Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Inteligence* 12 (629): 639.

Peterson, D.L. and Running, S. W. (1989) Applications in forest science and management. In: *Theory and Applications of Optical Remote Sensing* (ed Asrar,G.), Wiley, New York.

Piaget, J. (1969) *The Mechanisms of Perception*. Rutledge and Kegan Paul, London.

Pinilla, C. (1995) *Elementos de Teledetección*. Ra-Ma, Madrid.

Prigogine, I. (1962) *Introduction to Non-equilibrium Thermodynamics*. Wiley Interscience, New York.

Proisy, C., Mougin, E., Dufrêne, E., and Le Dantec, V. (1999) Monitoring seasonal changes of a mixed temperate forest using ERS SAR observations. *IEEE Transactions on Geosciences and Remote Sensing* 38 (1): 540-552.

Quegan, S., Le Toan, T., Yu, J. J., Ribbes, F., and Floury, N. (2000) Multitemporal ERS SAR analysis applied to forest mapping. *IEEE Transactions on Geosciences and Remote Sensing* 38 (2): 741-753.

Reed, B.C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., and Ohlen, D. O. (1994) Measuring phenological variability from satellite imagery. *Journal of Vegetation Science* 5: 703-714.

Richards, J.A. (1993) *Remote Sensing Digital Image Analysis: an Introduction*. Springer-Verlag.

Rissanen, J. (1978) Modeling by shortest data description. *Automatica* 14: 465-471.

Rosch, E. (1978) Principles of categorization. In: *Cognition and Categorization* (eds Rosch,E. and Lloyd,B.B.), Erlbaum.

Rowe, J.S. (1961) The Level-of-Integration Concept and Ecology. *Ecology* 42 (2): 420-427.

Roy, B. (1991) The outranking approach and the foundations of ELECTRE methods. *Theory and Decision* 31: 49-73.

Ruiz de la Torre,J. (1990) *Mapa Forestal de España escala 1:200.000. Memoria General*. ICONA, Madrid.

Salthe, S.N. (1985) *Evolving Hierarchical Systems: Their Structure and Representation*. Columbia University Press.

Salthe, S.N. (1993) *Development and Evolution: Complexity and Change in Biology*. MIT Press.

Saura, S. (2001) *Influencia de la escala en la configuración del paisaje: estudio mediante un nuevo método de simulación espacial, imágenes de satélite y cartografías temáticas.* Ph.D. Thesis, Universidad Politécnica de Madrid.

Schiewe, J., Tufte, L., and Ehlers, M. (2001) Potential and problems of multi-scale segmentation methods in remote sensing. *Geographische Informationssysteme* 6 (34): 39.

Schowengerdt,R.A. (1997) *Remote Sensing, models and methods for image processing*, Second edn. Academic Press, London.

Schrödinger,E. (1944) *What is Life?* Cambridge University Press.

Schwartzman,D. (1999) *Life, Temperature, and the Earth: The Self-Organizing Biosphere*. Columbia University Press.

Shannon, C. and Weaver, W. (1949) *The Mathematical Theory of Communication*. University of Illinois Press.

Short, N.M. (2002) *The Remote Sensing Tutorial*. NASA/Goddard Space Flight Center. http://rst.gsfc.nasa.gov/

Shuurman, N. (1999) Critical GIS: Theorizing and Emerging Science. In: *Cartographica Monograph 53* pp. 51-67. University of Toronto Press.

Sinton, D. (1979) The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. In: *Spatial Semantics: Understanding and Interacting with Map Data.* Proceedings, First International Study Symposium on Topological Data Structures for Geographic Information Systems 7: 1-17.

Smith, B. (1995) The structures of common-sense world. *Acta Philosophica Fennica* 58: 290-317.

Smith, B. (2000) Truth and the Visual Field. In: *Naturalizing Phenomenology. Issues in Contemporary Phenomenology and Cognitive Science* (eds Petitot,J., Varela,F.J., Pachoud,B., and Roy,J.M.), pp. 317-329. Stanford University Press.

Smith, B. (2001) Fiat Objects. *Topoi* 20 (2): 131-148.

Smith, B. and Brogaard, B. (2000) Quantum Mereotopology. In: *Spatial and Temporal Granularity* pp. 25-31. AAAI Press,

Smith, B. and Mark, D. M. (1998) Ontology and Geographic Kinds. 308-320. International Geographical Union. Proceedings. 8th International Symposium on Spatial Data Handling (SDH'98).

Smith, B. and Mark, D. M. (2002) Geographic Categories: An Ontological Investigation. *International Journal of Geographic Information Science* 15 (7): in press.

Smith, B. and Varzi, A. C. (2000) Fiat and bona fide boundaries. *Philosophy and Phenomenological Research* 60 (2): 401-420.

Smits, P.C. and Annoni, A. (1999) Towards operational knowledge-based remote-sensing image analysis. *Pattern Recognition Letters* 20 (11): 1415-1422.

Smits, P.C., Dellepiane, S. G., and Schowengerdt, R. A. (1999) Quality assessment of image classification algorithms for landcover mapping: a review and a proposal for a cost-based approach. *International Journal of Remote Sensing* 20 (8): 1461-1486.

Stehman, S. and Czaplewski, L. (1998) Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment* 64: 331-344.

Steinhaus, H. (1960) *Mathematical Snapshots*. Oxford University Press, London.

Stohlgren, T.J., Kalkhan, M. A., Chong, G. W., and Schell, L. D. (1997a) Multi-scale sampling of plant diversity: effects of the minimum mapping unit. *Ecological Applications* 7: 1064-1074.

Stoney, W. E. (1997) Land sensing satellites in the year 2000. Singapore, IEEE. IGARSS '97.

Stonier, T. (1990) *Information and the Internal Structure of the Universe*. Springer-Verlag, london.

Stonier, T. (1996) Information as a basic property of the universe. *Biosystems* 38: 135-140.

Strahler, A.H., Woodcock C., and Smith, J. A. (1986) On the nature of models in Remote Sensing. *Remote Sensing of Environment* 20: 121-139.

Swenson, R. (1989) Emergent Attractors and the Law of Maximum Entropy Production: Foundations to a Theory of General Evolution. International Federation for Systems Research . Pergamon Press.

Swenson, R. and Turvey, M. T. (1991) Thermodynamic reasons for Perception-Action Cycles. *Ecological Psychology* 3 (4): 317-348.

Talmy, L. (1996) The Windowing of Attention in Language. In: *Essays in Semantics and Pragmatics* pp. 235-287. John Benjamins. P.

Testa, B. and Kier, B. (2002) Emergence and Dissolvence in the Self-organisation of Complex Systems. *Entropy* 2: 1-25.

The Nature Conservancy (1994) Standardized National Vegetation Classification System. NBS/NPS Vegetation Mapping Program.

Thom, R. (1975) *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*. Benjamin-Cummings Publishing.

Thom, R. (1988) *Esquisse d'une sémiophysique*. Paris: InterEditions.

Thomasson, A.L. (2001) Geographic objects and the science of Geography. *Topoi* 20: 149-159.

Thrall, G.I. (1995) The stages of GIS reasoning. *Geo Info Systems* 5 (2): 46-51.

Ton, J., Sticklen, J., and Jain, A. K. (1991) Knowledge-based segmentation of Landsat images. *IEEE Transactions on Geoscience and Remote Sensing* 29 (2): 222-231.

Townshend, J.R.G. and Justice, C. O. (1988) Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations. *International Journal of Remote Sensing* 9 (2): 187-236.

Tönjes, R., Growe, S., Bückner, J., and Liedtke, C. E. (1999) Knowledge Based Interpretation of Remote Sensing Images Using Semantic Nets. *Photogrammetric Engineering and Remote Sensing* 65 (7): 811-821.

Tversky, B. and Hemenway, K. (1984) Objects, Parts and Categories. *Journal of Experimental Psychology: General* 113: 169-193.

Ulaby, F.T. and Dobson, M. C. (1989) *Handbook of Radar Scattering Statistics for Terrain*. Artech House, Norwood, MA.

van Fraasen, B. (1966) Singular Terms, Truth Value Gaps, and Free Logic. *Journal of Philosophy* 63: 481-495.

Varzi, A. (2001c) Vagueness in Geography. *Philosophy and Geography* 4 (1): 49-65.

Villaescusa, R., Vallejo, R., and De la Cita, J. (2001) Actualización del Mapa Forestal de España. III Congreso Nacional Forestal : 153-158. Granada, Junta de Andalucía.

Vincent, L. and Soille, P. (1991) Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (6): 583-598.

Visvalingam, M. (1991) Areal units and the linking of data: some conceptual issues. In: *Spatial Analysis and Spatial Policy using Geographic Information Systems* (ed Worrall,L.), pp. 12-37. Belhaven Press, von Bertalanffy,L. (1950) An outline of General Systems Theory. *British Journal for the Philosophy of Science* 1 (139): 164.

Wagner, W., Vietmeier, J., and Schmullius, C. (2000) Information content of ERS SAR interferometric products for forest classification in SIBERIA: a case study over the Bolshemurtinskii forest enterprise. IGARSS 2000.

Wang, D.C., Vagnucci, A. H., and Li, C. C. (1981) Gradient Inverse Weighted Smoothing Scheme and the Evaluation of its Performance. *Computer Graphics and Image Processing* 15: 167-181.

Weber, K.T. (2001) A method to incorporate phenology into land cover change analysis. *Journal of Vegetation Science* 54: A1-A7.

Weickert, J. (1998) Fast segmentation methods based on partial differential equations and the watershed transformation. In: *Mustererkennung 1998* (eds Levi,P., Ahlers,R.J., May,F., and Schanz,M.), Springer, Berlin.

Weickert, J. (1999) Nonlinear diffusion filtering. In: *Handbook on Computer Vision and Applications, Vol. 2: Signal Processing and Pattern Recognition* (eds Jähne,B., Haußecker,H., and Geißler,P.), pp. 423-450. Academic Press, San Diego.

Weickert, J. and Benhamouda, B. (1997b) A semidiscrete nonlinear scale-space theory and its relation to the Perona-Malik paradox. In: *Advances in computer vision* (eds Solina,F., Kropatsch,W.G., Klette,R., and Bajcsy,R.), pp. 1-10. Springer, Wien.

Wesolkowski, S. (1999) *Color Image Edge Detection and Segmentation: A Comparison of the Vector Angle and the Euclidean Distance Color Similarity Measures.* Master Sc, University of Waterloo, Canada

West,D.B. (2000) *Introduction to Graph Theory*, 2 edn. Prentice-Hall.

Whittaker, R.H. (1962) Classification of natural communities. *The Botanical Review* 28 (1)

Wiener N. (1948) *Cybernetics or control and communication in the animal and machine*. MIT Press.

Wiens, J.A. (1976) Population responses to patchy environments. *Annu.Rev.Ecol.Sys.* 7: 81-120.

Wiens,J.A. (1989) Spatial scaling in ecology. *Functional Ecology* 3: 385-397.

Wiens,J.A. (1995) Landscape mosaics and ecological theory. In: *Mosaic Landscapes and Ecological Processes* (eds Hansson,L., Fahrig,L., and Merriam,G.), pp. 1-26. Chapman and Hall, London.

Witkin, A. P. (1983) Scale-space filtering. Proc.8th Int.Joint Conf.AI : 1019-1022. Karlsruhe, Germany.

Wittgenstein, L. (1999) *Philosophical Investigations*. Prentice-Hall.

Wolfram, S. (1986) Approaches to Complexity Engineering. *Physica D* 22: 385-399.

Wong, D. and Amrhein, C. (1996) Research on the MAUP: Old wine in a new bottle or real breakthrough? *Geographical Systems* 3: 73-76.

Woodcock, C.E. and Harvard V.J (1992) Nested-hierarchical scene models and image segmentation. *International Journal of Remote Sensing* 13 (16): 3167-3187.

Woodcock, C.E. and Strahler, A. H. (1987) The factor of scale in remote sensing. *Remote Sensing of Environment* 21: 311-332.

Wu,J. (1999) Hierarchy and scaling: Extrapolating information along a scaling ladder. *Canadian Journal of Remote Sensing* 25 (4): 367-380.

Wu, J. and Loucks, O. L. (1995) From balance-of-nature to hierarchical patch dynamics: a paradigm shift in ecology. 70:439-466. *Quarterly Review of Biology* 70: 439-466.

Young, L.B. (1993) *The Unfinished Universe*. Oxford University Press, Zadeh, L. (1965) Fuzzy sets. *Information and Control* 8 (3): 338-353.

Zhang, Y.J. (1996) A survey on evaluation methods for image segmentation. *Pattern Recognition* 29 (8): 1335-1346.

APPENDIX 1. The hierarchical organisation of the Universe

The world can be conceived a complex hierarchy of nested levels. Every real entity (an object) is made up of parts (another objects) that are obviously smaller than the whole. What is a part of an object at some level of the hierarchy becomes an object itself at a lower level and vice versa[1]. We can go down from the whole biosphere to biomes, every one made up of different kinds of ecosystems, like e.g. a forested area that may be divided into hardwood and conifer forests, and the latter into pine and spruce stands that are made up of individual trees. We might be interested in going further down and look at the branches, leaves, parenquimal cells or chloroplasts of an individual tree, either further up and consider the Earth, the Solar System, the Milky Way, the Local Group of galaxies up to the whole Universe.

The hierarchy is not only structural but also functional, since these chloroplasts of that needle are reflecting green light that summed to the other needles of that spruce, and to the others of neighbouring trees and to the whole stand and forest and landscape yield a local value of albedo[2] that in turn contribute to the entropy production of our planet. The hierarchic organisation allows this reflective property to be measured at each level without having to know what interactions (mostly non-linear) between the objects in the lower levels led to the observed measurement. The study of complex systems showing this organization is addressed by a new branch of General Systems Theory (von Bertalanffy 1950) called Hierarchy Theory (Pattee 1973;Ahl & Allen 1996), that can help to understand ecological complexity and scale. The following two paragraphs explain briefly the major points of this theory. They have been condensed from an interesting article (1999c) by the American ecologist Jianguo Wu, who has integrated these concepts into a new paradigm in ecology, the hierarchical patch dynamics (Wu & Loucks 1995).

A hierarchical system has both vertical structure that is composed of levels and horizontal structure that consists of holons. Hierarchical levels are separated, fundamentally, by characteristically different process rates. The boundaries between levels and holons are the places showing the highest variability in the strength of interactions. In hierarchical systems, higher levels are characterized by slower and larger entities (or low-frequency events)

[1] For this reason, Arthur Koestler (1967b) pointed out that the word hierarchy in this context should be replaced by *holarchy*, since they are composed of holons (wholes that are simultaneously parts of other wholes).

[2] the fraction of incident radiation (as light) that is reflected by a surface or body (as the moon or a cloud).

whereas lower levels by faster and smaller entities (or high-frequency events). The relationship between two adjacent levels is asymmetric: the upper level exerts constraints (e.g., as boundary conditions) to the lower level, whereas the lower provides initiating conditions to the upper. On the other hand, the relationship between subsystems (holons) at each level are symmetric, and can be distinguished by interacting more strongly or more frequently within than between them. For example, the strength of interactions between subatomic components is stronger than that between atoms which is in turn stronger than that between molecules. The same can be said about an ecological hierarchy such as the nested hierarchy of tree-stand-forest-landscape. Therefore, it is the variability in the strength of interactions between levels and among holons (patches) that defines the locations of boundaries, and it is the relatively high degree of interactions among components that gives rise to the apparent identity and integrity of holons as well as systems.

These characteristics of hierarchical structure can be explained by virtue of "loose vertical coupling", permitting the distinction between levels, and "loose horizontal coupling", allowing the separation between holons at each level. Strictly speaking, complete decomposability only occurs when coupling between components becomes zero, which seems a trivial case because, by definition, a system is composed of interacting parts. Thus, hierarchical complex systems are only nearly decomposable. However, near-decomposability seems to underline the plausibility and success of seemingly independent and partial studies of nature crossing different hierarchical levels, ranging from elementary particles to the cosmos. Hierarchy theory suggests that when one studies a phenomenon at a particular hierarchical level (the focal level), the mechanistic understanding comes from the next lower level, whereas the significance of that phenomenon can only be revealed at the next higher level. Thus, three adjacent levels or scales usually are necessary and adequate for understanding most of the behavior of ecological systems. This triadic structure and the nearly decomposable nature of complex systems provide a key to their simplification and manageability. But now, why are complex systems, and ultimately the Universe, organised like this?

Hierarchical organisation is a consequence of the natural tendency of matter towards integration. The most conspicuous sign of this tendency is gravitation, the primary force giving structure to Universe. Another natural tendency of the Universe, expressed by the Second Law of Thermodynamics, is the search for global equilibrium, the dissipation of energy, i.e. the destruction of all gradients from field potentials (non-uniform distributions of energy). Gravitation has formed clumps of matter (e.g. planets) that following Prigogine (1962) can be regarded as dissipative structures (i.e. receiving low entropy[1] energy from stars and radiating it back as high-entropy heat) far from equilibrium (i.e. undergoing continuous change) but maintaining a persistent steady-state that is continuously fluctuating through irreversible processes. These dissipative structures contribute to the *heat death* (the sought equilibrium) of the Universe.

On the other hand gravitation creates field potentials, forcing a non-uniform distribution of matter (which in the end is 'delayed' energy) throughout Space, so there is an apparent contradiction in the way Universe seeks equilibrium. A solution for this paradox is the *maximum entropy production* principle (MEP) suggested by Rod Swenson (1989c). Actually, MEP was first proposed by Paltridge (1978a) in the context of Meteorology. He suggested that the long-term mean state of the Earth's climate may represent a maximum rate of entropy increase by the turbulent heat transport in the atmosphere and oceans. Swenson goes beyond thermal convection and proposes a guiding principle for a future Theory of general evolution.

According to the MEP principle, dissipative systems (sets of dissipative structures hierarchically nested) evolve in the direction of the most rapidly dissipative states –those faster in minimising field potentials, that is, in maximising the entropy- given the conditions to which they are constrained. Since the more ordered a system is, the more resources must consume (therefore in cases where the energy input is fixed, it must enhance its energetic efficiency) in order to maintain its structure and functioning, it seems that the best way to produce entropy as fast as possible should be by producing order as fast as possible, 'because order produces entropy faster than disorder' (Swenson & Turvey 1991).

This would explain the recurrent self-organising patterns found in (not only living) Nature, for they are -as a consequence of MEP- merely opportunistic: they occur as soon as they get the chance. The biosphere is one of these self-organising systems following MEP, growing both horizontally (increasing the amount of biomass) and vertically (creating higher levels of organisation, i.e., increasing complexity). This view is supported by the research of biologist

[1] Entropy here relates inversely to the capacity of transforming that energy into useful work (i.e. a measure of the inefficiency of that process), giving an idea of the 'quality ' of energy. More generally, entropy is a measure for the amount of disorder or 'chaos' in a system. The greater the entropy, the more uniformly distributed the system is, i.e. the lower their field potentials are, and so the less work it can perform. Therefore both concepts are identical.

David Schwartzman (1999d). He hypothesized that the cooling of the Earth's surface since the origin of life is a result of the progressive self-organization of the biosphere that has affected atmospheric composition (see figure 1-1). He modelled the net entropic flow from the Earth surface since geological times and found that the trend in entropy flux is consistent with MEP. Another supporting evidence comes from Holbo and Luvall (1989d). They measured surface temperatures of various ecosystems using the Thermal Infrared Multispectral Scanner (TIMS), and found that when other variables are constant, the more developed the ecosystem the colder its surface temperature.

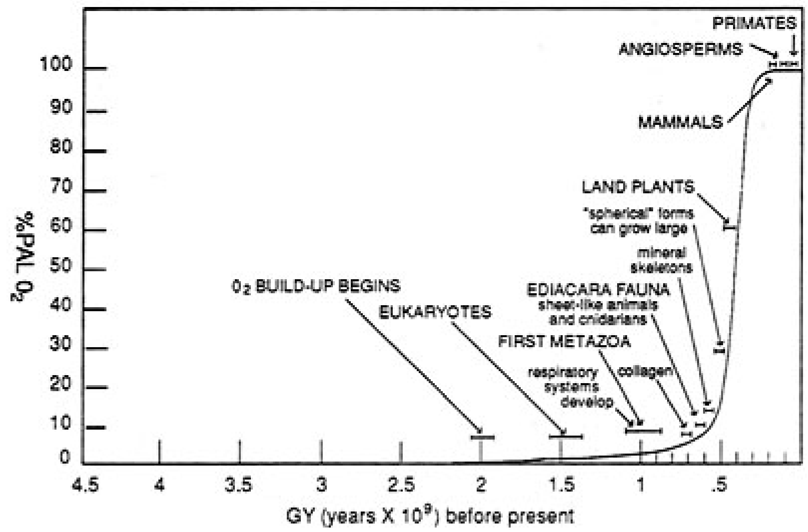


Figure A1-1 Buildup of atmospheric O2 in geological time (PAL is present atmospheric level). From Swenson (1989)

So far we have briefly described the hierarchical organisation of the Universe and suggested a final cause for its existence, but have not explained the synergetic mechanism producing higher levels of complexity. Higher order involves a number of individual systems or units working together to produce a new and larger whole with properties that trascend those of the units (Young 1993). This coupling results in a reduction in the number of probable states[1] of the constitutive units and/or in an alteration of the relative probability of these states. This phenomenon , coined by Testa and Kier (2002) *dissolvence*, allows the *emergence* of new properties more efficient in terms of entropy production than the ones lost. The increased dissipative efficiency of the whole enhance its adaptability to environmental changes through what Bronowski (1970) called *stratified stability*, that is, a sort of buffering effect between levels. The stability is achieved because the lower levels have a more intense dissipative

[1] These states are generated by fluctuations that change the structure and function of the system. Structure, function and fluctuation are therefore essential interdependent attributes of adaptative complex systems.

activity (higher metabolic rates), so they reach thermodynamic equilibrium (death) more rapidly, enabling an almost adiabatic interaction (basically consisting of informational exchanges) across levels (Salthe 1993).

Testa and Kier (2000) give several examples of dissolvence. At the molecular level, electron sharing among the atoms of a molecule produce the dissolvence of some atomic properties together with the emergence of some molecular ones, the latter resulting from the coherent collective behaviour of the atoms involved. At the biological level, each cell nested into a pluricellular organism is an open system controlled by higher-level systems (tissue, organ and organism) via a permanent flux of matter, energy and information. By *dissolving* themselves into a greater entity, they have lost the independence of prokaryotic cells (e.g. bacteria), but in turn the new entity can perform a number of new powerful functions not available for the unicellular organism. Finally, at the social level, the most conspicuous example of dissolvence/emergence, apart from human societies, is found in social insects (e.g. bees, ants and termites). These species form highly integrated colonies which closely resemble a pluricellular organism. In a colony, differentiation, behaviour and even death of individuals are controlled by chemical signals emanating from the queen. The specialised individuals cannot survive for long away from their colony, but in turn the emergent properties of the colony involve *inter alia* an optimised use and distribution of limited resources, homeostatic conditions in the its interior, and improved defence against predators and competitors.

In short, the creation of higher levels of order is a built-in tendency of our evolving Universe that have led to the hierarchical structure we observe. Higher complexity in a given system arises as a result of dissolvence/emergence processes, and the consequence is that the system's behaviour becomes more and more autonomous relative to its material basis (Hoffmeyer 1997). Although the mechanisms underlying this evolution are not yet well known, it seems to me, from the limited survey I have made, that we are close to unveil the mystery of consciousness without having to resort to religious or mystic explanations. Such an achievement (i.e., the dissolvence of the ancient mind/body dualism) would represent an extraordinary scientific revolution, enabling the construction of an image of the Universe with the observer inside it. This prospect is so promising that, albeit not directly related to the issues addressed in this thesis, I considered I should at least mention it. In any case, ‘hierarchical structure provides a strong theoretical basis for explaining the problem of scale and for developing approaches to it' (Wu & Loucks 1995). This is the ultimate reason for the inclusion of this appendix in the thesis.

APPENDIX 2. The concept of information

The vast majority of computing, telecommunications, remote sensing, etc literature uses the term *information* as basic concept which is however left unexplained, i.e. the concept of information is treated as a *black box* whose content need not to be questioned. In many texts, data are seen as huge collections of numbers that only become information after having been processed or evaluated for a particular purpose, whereas knowledge is accumulated information that has been verified and made usable prior to storage. However, these words are often used as if they were synonymous, fuelling the confusion.

To worsen the situation, the only Information Theory so far accepted has nothing to do with information, as his author, Claude Shannon (1949), did tacitly admit[1]. The layman's notion of information –a piece of knowledge endowed with meaning- is in contrast to Shannon's definition as a purely quantitative measure of communicative exchanges through a noisy channel. Actually Shannon's is more a signal transmission theory than an information one, since the problems of interpreting signals into a meaningful message are left outside his theory.

A better approach towards a concept of information comes from the fields of non-equilibrium thermodynamics, semiotics[2] and cybernetics[3]. Following Wiener's (1948) view, the amount of information of a system is linearly related to its degree of organisation, i.e. to the amount of order, where *order* is any non-random pattern of matter or energy existing in the system. With this premise and Schrödinger's (1944) consideration of order as the inverse of disorder (so that any process that gains information loses entropy), biologist Tom Stonier (1990) derived an equation in which Boltzmann's entropy is the multiplicative inverse of the information content of a system, so that one entropy unit has roughly 10^{23} bits/mole. This puts information

[1] By saying that 'the semantics aspect of communication are irrelevant to the engineering problem' [ibid, p3]. Due to this acknowledgement, Shannon was unsure about how to call his measure. He first coined it *uncertainty*, but later he adopted the term *entropy* following Von Neumann's advice that this term would promote discussion of his theory "because nobody would understand its meaning" (source: M. Tribus, 1971: Energy and Information. In Scientific American,3:179-188).

[2] Semiotics is concerned with everything that can be taken as a sign, where 'sign' refers to anything which stands for something else. Founded in the first third of 20th century by linguist Ferdinand de Saussure and philosopher Charles S. Peirce, it studies how meanings are made: as such, it is concerned not only with communication but also with the construction of reality (Chandler 2001).

[3] Cybernetics, founded by Norbert Wiener in the half of last century, studies physical systems whose parts interact (communicate) in a controlled way as to promote a particular configuration or goal-state of the system. This drive or purposiveness can be internal (teleonomic) or externally imposed (teleologic). Wiener derived the name from the Greek *kybernetes*, steersman.

as a basic property of the Universe, the way mass and energy are: just as mass is a property of a system having matter, and energy is a property of a system whose constituent particles move, so is information a property of a system showing organisation (Stonier 1996).

Although Stonier's calculations may be wrong (it does not seem so easy to reduce to physical order the relational aspects of information), the basic idea is appealing, since it is clear that there is more than matter and motion (energy) in the universe: there is order, and information is the magnitude emanating from it. If this view is accepted, it can be said that information, like energy, may appear in many forms, and will remain *latent* until some entity makes use of it. Just as potential energy of water flowing down a river cannot be converted into useful work until e.g. a hydroelectric station is built, so is latent information coming from the environment not seized until a study is carried out to decide where the station is to be placed. **So if energy is the capacity to do work, (functional) information is the capacity to control this capacity, i.e. the capacity to regulate the acquisition, storage and consumption of matter/energy** (Corning 2001).

This definition is obviously dependent on the cybernetic system who exerts the control (the *user*), and its relationship (the *context*) with the *environment* (the source of information), so that latent information may be gathered and used in many ways. The meaning of functional information is given by its impact on the current or future behaviour of the user (therefore this meaning may change from user to user, an even for the same user after some time, as result of feedback processes), and its value depends on the consequences (benefits or damage) of actions (or inaction) taken upon this information. As noted by Corning (2001), the former definition enables an economic appraisal of information. If the efficiency (benefit-cost ratio) is low, it is likely that it will not be gathered, remaining in the realm of latency. If on the other hand the potential benefits are high and/or the cost of that information is reduced, not only will it be collected once, but it will be renewed periodically as to maintain the efficiency high. However, there is still an unanswered question: what is the link between order and latent information?

Order in a system implies an underlying structure that is manifested through patterns, and patterns are formed by differences between components and/or by differences between the system as a whole and its surroundings. If *difference* is any relation of non-identity between physical entities or their properties (size, shape, colour, temperature, etc), then **latent**

information can be defined as any detectable difference (Chmielecki 1998). For this difference to become functional information, it has first to be *detected* by another system –the so called *user*- via a sense organ or *epistemon* (Barham 1996), and *interpreted* together with other differences in such a way as to represent the sensed structure using the salient aspects of it, that is, as to construct an internal model of the system under observation, a model that gives *sense* to the detected differences. Thus latent information has no meaning, it arises from interpretation, in what can be called *semiosis*, i.e. the process of meaning construction. Semiosis is above all an economical process that follows the Minimum Description Length (MDL)[1] principle: from the vast set of possible differences, what gets into the model is only the small subset of relevant ones.

I will call this representation endowed with meaning *structural information*, which paraphrasing Gregory Bateson[2] (1972), may be defined as *a selection of sensed differences that makes sense*. Note that for structural information to be communicated, it has to be materialized into another structure, namely a formal representation *–model-* of the internal one. Thus, **communicable structural information** is rather symbolic, for it may be anything that stands for another thing. The link between both things (the *code*) is established by the user, and this code must be known by the intended receiver of the information. Thus a practical definition of this kind of information is: **a formal representation of a piece of reality whose meaning can be agreed by a community of users**.

Using Korzybski's (1933) metaphor, some **properties of communicable structural information** can be listed: it is **symbolic** (the map is not the territory for which it stands), **isomorphic** (the features of the map must have a fixed correspondence with the ones of the territory), **incomplete** (no map includes all the features of the territory), **subjective** (no map exists free of some kind of contamination from the map-maker), and **recursive** (every user makes its own *map* out of a map, which is not identical to the original one). Note that some of the differences portrayed in the model may be virtual (not detected but imposed by the observer for *gestalt* reasons –proximity, similarity, symmetry, closure, etc), as long as they are needed to give a meaningful form to the represented structure. Note also that since

[1] Proposed by Rissanen (1978b), MDL states that the best model to explain (i.e. give meaning to) a set of data is the one which minimizes the sum of i) the length (e.g. in bits) of the description of the model, and ii) the length of data, when encoded with the help of the model. MDL is th reason behind raster to vector conversion.

[2] In a brilliant lecture entitled '*Form, Substance and Difference*', Bateson argued that from the infinite number of differences that can be found between and within objects, we select a very limited number, which become information, so that the elementary unit of information is "*a difference that makes a difference*".

representational isomorphisms are transitive, information may be transduced (e.g. from sonic waves to electromagnetic ones) and encoded (e.g. into electric pulses or into bits) while still representing the same thing (Chmielecki , 1998).

It is the isomorphic property of structural information that makes it useful, for if there were no fixed correspondence between the representation and the environment, the user could not foresee the results of his actions and therefore would show a chaotic behaviour that would point him towards thermodynamical equilibrium. For example, if the chameleon's vision would not enable a correct aiming of his tongue, he would starve. So in the end, **'the meaning of information is the prediction of the success of functional action'** (Barham, 1996). If in addition to *epistemons*, the user has some capacity to, on the one hand, store, retrieve or erase (that is, memory) previously perceived structural information, and on the other, bind through some relational framework (that is, logics) that information, then it can reuse and even *create*, i.e. *infer*, new information that will enhance the reliability of the prediction.

With this built-in capability (the *know-how*), the user can compare the current situation to past ones, allowing inferences about properties of interest (that is, about what the *environment* can afford for the *user*, i.e., *affordances* in Gibson's (1979) terminology) not directly detected. If the inference is wrong, the action taken upon it is likely to fail, therefore the meaning of the pieces of information involved will be reassessed. If the action is successful, it will support the inference made, and this success will be knitted to the information in the user's memory. I would like to conclude this reasoning by proposing another odd definition: **knowledge is any story of success**, where *story* is a set of inter-related frames of structural information, and *success* is the positive verification of the story through the result of past decisions (made possibly by other users) based upon that *story*.

To summarize, information comes ultimately from physical order. Order implies differences, that is, non uniform distributions of matter/energy. Certain systems having sensors can take advantage from these differences for cognitive purposes, as long as they can detect them. From all the set of detectable differences (**latent information**), only the relevant ones are used by those systems to construct a handy representation (**structural information**) of the sensed scene, that has to be formalised into a model in order to be communicated. If the system has cybernetic –communication and control- capabilities, and the model is taken into account in its decision-making process, then it becomes **functional information**. The meaning of that information is the prediction of the success of the selected action, and its value depends on the consequences of that action. If the result of the action is the one expected, then this information is stored as (reusable) **knowledge**.

APPENDIX 3. Resolution-limited representations of geographic space

Class attributes may be modelled as regionalised variables, or *geographic fields* (defined in 2.2.21), whose value is dependent on location: $A(x,y,t)=a$, where a is the value taken by attribute A at time t in a geographic point of latitude x and longitude y. Similarly, a categorical representation (thematic map) can be modelled as a regionalised discrete variable returning the label at each point. On the other hand, every formal representation is finite by necessity, and so is its level of detail (a 1:1 map with no generalisation would not be a map but the territory itself). Therefore the boundaries represented in a thematic map are finite approximations of the boundaries at infinite resolution (reality) of the objects of interest. Confronted to the problem of representing boundaries at different resolutions, Bud Bruegger developed in his Ph.D thesis (1994) a new spatial theory based on resolution disks associated to euclidean points, where the diameter of the disks is equated to the resolution of the representation, likewise the GIFOV in remote sensing.

The surface of a disk crossed by an infinitely thin boundary separating objects of different class will consists of a mixture, where the proportion of each class can be precisely measured in infinite-resolution space. Resolution limited-space is constructed by a mapping that returns the mixture percentage found in a disk centred at each point. Then the classes are reinterpreted in resolution-limited space as *mixture classes*, where each mixture class is defined by its predominant mixture component. This predominance is expressed by a minimal percentage (fig. 2-4) which is called the *level of homogeneity*. Euclidean points whose associate disk contains no class above this level, are part of the *transition zone* (TZ) that separate objects. The width of TZ will depend on the configuration of objects in geographic reality, but also on the resolution (disk size) relative to the mean size of objects, and on the chosen level of homogeneity. Finally, this resolution-limited representation could be transformed to vector format by using the medial axis of TZ as the polygon boundaries (fig. 2-5). TZ could therefore constitute an ancillary polyline layer where each arc would have the width of TZ as its attribute. In this way, a formal account of boundary placement uncertainty due to the limited resolution of the representation could be added to vector layers.

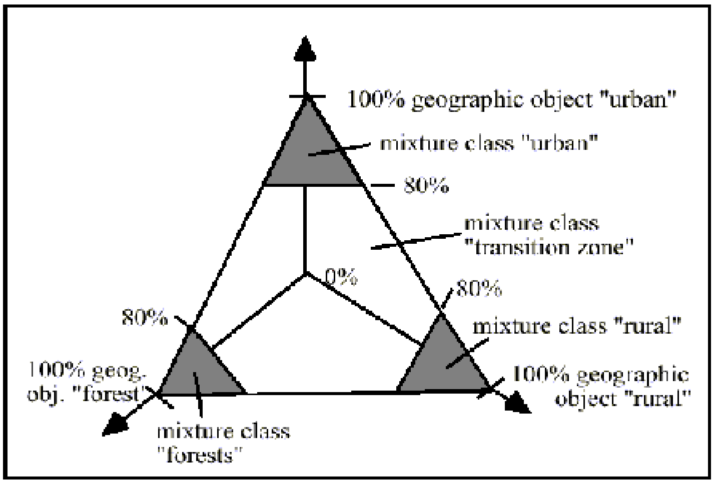


Figure A3-1. Mixture classes defined by a level of homogeneity of 80%. From Bruegger (1994)

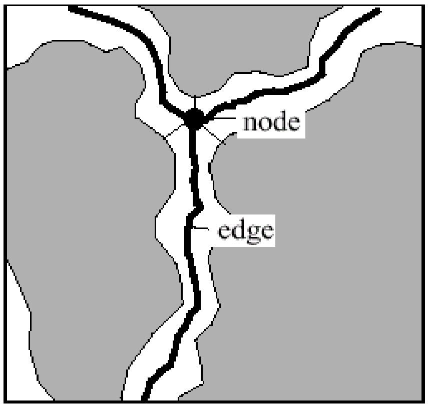


Figure A3-2. Medial axis of the transition zone. From Bruegger (1994)

The former idealised process can also be applied to change from a already resolution-limited representation to a coarser one, and indeed this is the main practical application of Bruegger's theory. Note that in this case, the uncertainty due to imprecise knowledge about i) the mixture distribution, and ii) the already effected approximation of boundaries, cannot be captured in the TZ layer.

To end up, Bruegger's spatial theory has served to this thesis as a source of inspiration to conceptualise the 'measurement disk' notion, by transferring the concept of 'resolution disk' from the realm of representation to the one of mensuration. This is why it has been briefly outlined, though it is not used explicitly in the model of geographic reality proposed here.

APPENDIX 4. The Forest Map of Spain (MFE)

In this appendix, the actual map that is used as a running example to illustrate the multi-tiered model of this thesis is briefly described. The Forest Map of Spain (MFE, Mapa Forestal de España) consists of a set of ninety two 1:200,000 sheets covering the whole Spanish territory. Each sheet portrays a rectangle of some 115x70 km^2 in which, using a topographic map as background, a network of polygons involving internal homogeneity in terms of floristic composition and physiognomic structure is displayed. The focus of MFE is natural and semi-natural vegetation, including not only forest but also shrubland, grassland and sparsely vegetated wild areas.

MFE was initiated in 1986 by the Directorate of Nature Conservation (DGCONA) of the Spanish Government with the aim of compiling an up-to-date vegetation map to use it as a natural resource inventory and as baseline to land use planning and environmental monitoring. A previous compilation did exist, but it was out of print, outdated (it was published in 1966), and less detailed (the scale was1:400,000 and only one species was consigned in each polygon). The new compilation (MFE-2C) was entrusted to Prof. Juan Ruiz de la Torre, chair of the Botany Department of the Forest Engineering School of Madrid. The methodology he devised (Ruiz de la Torre 1990) is similar to Kuchler's (1967a) comprehensive method for vegetation mapping.

The first stage of the mapping procedure involved the interpretation of panchromatic stereo pairs from the national 1:30,000 photogrammetric flight of 1985. Afterwards, the photointerpreted polygon boundaries were transferred through optical rectification to a 1:50,000 cartographic base in UTM projection (the same used in 1:200,000 maps). Then each polygon was inspected on the ground by a field team consisting of two botanists, consuming as an average 0.5h per km^2 of area of interest. The homogeneity of each polygon was checked, and boundary corrections were made where appropriate. After correction, several physiognomic and floristic variables were recorded in field data sheets for each remaining polygon. Subsequently, the corrected boundaries were transferred to the 1:200,000 cartographic base, and a descriptive label , background colour and other signs were allocated to each polygon. Each final map sheet is accompanied by a text report describing the main features (climate, geology, vegetation, soils) of the represented territory. The report also includes a complete listing of all the polygons, in which the main understory species found in each polygon are cited together with other remarks.

During the eleven years that took MFE-2C to complete, the GIS market spread at an amazing pace, establishing an scenario not foreseen at the beginning of the project . MFE-2C was conceived as a visual spatial database about vegetation, intended for consultation by a wide audience. The situation now is that most users demand not only hardcopy visual information, but a vector layer with its associated database that can be overlaid in a GIS with other sources of information and analysed quantitatively. MFE was also intended as a means for allocating surfaces to the plot network of the 2nd National Forest Inventory (IFN2). However, it was soon realized that MFE-2C representational scale was insufficient for this objective, due in part to the considerable percentage of forested areas catalogued as mosaics (some 25%). With these limitations and the need of a new inventory (IFN3) in mind, DGCONA approved in 1997 the updating of MFE at a 1:50,000 working scale (MFE-50).

The updating project, which began in 2000, makes use of a dedicated GIS, called Dinaforest, that allows the interpreter to create a preliminary georeferenced vector layer on the fly. Dinaforest is able to ingest a set of contiguous 1m-resolution panchromatic orthophotos in order to generate a UTM referenced digital photo-mosaic. The orthophotos come from the 1:20.000 national photogrammetric flight of 1997-98. The mosaic is then displayed on a PC so that polygons can be directly digitised on the screen. The digitising screen has a series of ancillary windows where another data from the currently visualised area (each updating unit is a 26x19 km^2 rectangle coincident with a 1:50,000 topographic sheet) is available, as e.g. Landsat TM images, the raster representation of MFE-2C, data from the IFN2 plots, etc. In this way the interpreter is able to allocate a label to most of the updated polygons at the same time he/she draws the boundaries. Polygons where the assignment is unclear are marked for inspection on the field survey. The number of 'mosaic' polygons has been considerably reduced as a result of disaggregation, although they are unavoidable, *inter alia* because of the chosen MMU (2.25 ha for forest polygons and 6.25 ha for other types). In principle, MFE-50 will not be distributed in hardcopy, rather it will be available via the internet from the Natural Resources Database (BDN) of the Ministry of Environment, who plans to renew it every ten years. Further information on MFE-50 can be found in (Villaescusa, Vallejo, & De la Cita 2001).

The vegetation attributes reported in each MFE polygon are the following:

- **i)** floristic composition;
- **ii)** abundance (ground cover fraction) of the relevant (tree or shrub) species;
- **iii)** height interval in which the relevant species are;
- **iv)** horizontal pattern of distribution of the relevant species;
- **v)** pattern of distribution of the different facets that compound a mosaic polygon;
- **vi)** areal percentage occupied by each type of facet in a mosaic polygon;
- **vii)** origin of the vegetation (natural or cultural);
- **viii)** soil condition (only if abnormal, like dunes or marshes; otherwise not reported);
- **ix)** potential vegetation formation (broad climatic category); and
- **x)** maturity level of actual vegetation.

The floristic composition (listed in the text report) consists of a non-exhaustive list of vascular plant species, of interest but not necessarily woody, present in each polygon. In contrast, relevant species (appearing in the polygon's label) are those that conform the uppermost stratum, normally consisting of trees or shrubs. If this stratum is sparse (less than 35% cover), then the species of the next taller (shrub or bush) stratum are also relevant. For each polygon, up to four relevant species can be reported in the label.

The cover fraction of each relevant species is computed as a percentage of the total area of the polygon, unless it is a mosaic, for in this case the percentage refers only to the area occupied by the facets in which this species occur. In case of overlapping, the tallest cover takes precedence. If the overlap takes place in the uppermost stratum (e.g. a two dominant species forest with overlapping crowns), the overlap area is distributed proportionally between the involved species. If there are no trees nor shrubs, the overall vegetation cover is given without reporting its distribution by species.

There are five height intervals in which the relevant species can be located: trees (> 7m), shrubs (3-7 m), dwarf shrubs (1.5-3 m), bushes (0.5-1.5 m), herbs and dwarf bushes (5-50 cm), and moss and creeping plants (< 5cm). There are some exceptions for the tree category, which can be reached with only 5m for sparse woodlands of the Quercus genus. The height interval for each relevant species is estimated visually without direct measurement, based on the average height of the canopy (or the next tallest stratum, if the former is sparse) within the polygon.

Regarding the spatial distribution of each relevant species within the polygon, five categories are distinguished, which are reported only for trees or, in case there are no trees, shrubs: *uniform*, if the species is even and densely distributed; *geometric*, a special case of the former where the distribution follows a geometric pattern, like e.g. in plantations; *in coppices*; *in hedges or rows*; *scattered*; and *multiple*, when more than one of the former distributions occur within the polygon. In the case of mosaic polygons, two additional attributes are given. The first one refers to the spatial arrangement of facets within the mosaic: *irregular*, if facets are randomly distributed; *mixed-irregular*, alike the former, but some of the relevant species more abundant in some type of facet appear closely mixed with others in the matrix or in another type of facet; *aspect*, if the orientation of facets depends on aspect (sunny versus shady hillsides); *altitudinal*, if the facets are arranged along horizontal stripes, changing from mountain's foothill to top; and *dendriform*, in hilly arid landscapes where the vegetation in gullies is different than in hillsides and/or hilltops. The second attribute for mosaics, occupation, is facet-specific and is the areal percentage occupied by each type of facet occurring within the polygon. The sum of occupations within a polygon must be 100.

The last four attributes (origin, soil condition, climatic formation and maturity level) can be inferred from the formers and from the geographic location, so that they can be assumed to be filled at office, and therefore will not be considered in the following discussion. Finally, note that MFE is a descriptive map that uses no explicit classification scheme, although the selected thresholds in height and cover impose a division of vegetated areas which, together with the use of botanical taxonomy, conforms MFE's implicit classification. Therefore MFE bears a great deal of local information that is very rich but difficult to assimilate in a synoptic way. As a matter of fact, it is hard to find in a given sheet two polygons with the same value in all the former attributes. Nevertheless, such difficulty can be easily tackled by granting each polygon membership to some class of a (preferably hierarchical) standard classification, as e.g. the UNESCO or the CORINE ones. Actually something of the like is being done for MFE-50, where each polygon is assigned to one of 27 classes. In this way, adjacent polygons belonging to the same class can be merged, and hence a more general analysis can be carried out.

APPENDIX 5. List of Acronyms

ANN, Artificial Neural Network

AVHRR, Advanced Very High Resolution Radiometer

BD, Bit Depth

CCD, Charge Coupled Device

CF, Cover Fraction

COSP, Change Of Support Problem

CV, Coefficient of Variation

DEM, Digital Elevation Model

DGPS, Differential Global Positioning System

DN, Digital Number

ECHO, Extraction and Classification of Homogeneous Objects

EO, Earth Observation

EPS, Edge Preserving Smoothing

FNEA, Fractal Net Evolution Approach

GIFOV, Ground-projected Instantaneous Field Of View

GIS, Geographic Information System

GIWEPS, Gradient Inverse Weighted Edge-Preserving Smoothing

GPS, Global Positioning System

GSI, Ground Sampling Interval

HR, High Resolution

HRVIR, High Resolution Visible and Infrared sensor

IDL, Interactive Data Language

IFOV, Instantaneous Field Of View

I-model, Idealistic model

IMORM, Iterative Mutually Optimum Region Merging

ISPRS, International Society of Photogrammetry and Remote Sensing

LAI, Leaf Area Index

LEO, Low Earth Orbit

LIDAR, Light Detection and Ranging

LUT, Look Up Table

MAUP, Modifiable Areal Unit Problem

MDL, Minimum Description Length principle

MEP, Maximum Entropy Production principle

MFE, Forest Map of Spain

MMU, Minimum Mapping Unit

MODIS, MODerate resolution Imaging Spectroradiometer

MSS, Multi-Spectral Scanner

NASA, National AeroSpace Administration

NDVI, Normalised Difference Vegetation Index

nRMSE, normalized Root Mean Square Error

NVCS, National Vegetation Classification System

NVD, Normalised Vector Distance

OOA, Object Oriented Analysis

OOCIM, Object-Oriented Classification of RS Images for landcover Mapping

PC, either Personal Computer or Principal Component

PSF, Point Spread Function

RGB, Red, Green and Blue

R-model, Realistic model

RMSE, Root Mean Square Error

RS, Remote Sensing

SAR, Synthetic Aperture Radar

SCRM, Size Constrained Region Merging

SMA, Spectrometric Approach

SPOT, Système Pour la Observation de la Terre

TCF, Tree Cover Fraction

TIMS, Thermal Infrared Multispectral Scanner

TM, Thematic Mapper

TZ, Transition Zone

UAV, Unmanned Aerial Vehicle

UN, United Nations

UTM, Universal Transverse Mercator

VHR, Very High Resolution